Robotic vocalization training system for the auditory-impaired

M Kitani, T Hara, H Hanada, H Sawada

Department of Intelligent Mechanical Systems Engineering, Faculty of Engineering,
Kagawa University, 2217-20, Hayashi-cho, Takamatsu-city, Kagawa, 761-0369, JAPAN

s10d501@stmail.eng.kagawa-u.ac.jp, sawada@eng.kagawa-u.ac.jp

http://www.eng.kagawa-u.ac.jp/~sawada/index_e.html

ABSTRACT

The authors are developing a vocalization training system for the auditory-impaired using a talking robot. The training system mainly consists of a talking robot which has mechanical organs like a human. With an adaptive learning strategy using an auditory feedback control, the robot autonomously learns the vocalization, and then reproduces the speech articulation from inputted sounds. By employing the talking robot, the training is realized by two different approaches. One is a training based on the hardware demonstration, which shows the speech articulation by the robotic motions, and the other is a software-based training, which shows the phonetic characteristics of generated voices. Training experiments are being conducted in Kagawa Prefectural School for the Deaf, and significant results have been obtained. In the previous system, the speech learning algorithm of the robot was constructed by using a Self-organizing Neural Network (SONN), which consists of the combination of a Self-organizing Map (SOM) and a Neural Network (NN). However, improper maps were found in the results of the speech articulation learning. In this study, a new algorithm using two SOMs, called a dual-SOM, is introduced for the autonomous learning of the robotic articulations. Firstly, the construction of the training system is described together with the autonomous learning of robotic vocalization using the dual-SOM algorithm, and then the analysis of the speech training progress is presented based on the phoneme characteristics and the mechanical vocal articulations.

1. INTRODUCTION

A voice is the most important and effective method of verbal and nonverbal communications. Various vocal sounds are generated by the complex articulations of vocal organs such as lung, trachea, vocal cords, vocal tract, tongue and muscles. The airflow from the lung causes the vocal cords vibration and generates a source sound, then the sound is led to a vocal tract to work as a sound filter as to form the spectrum envelope of a particular sound. The voice is at the same time transmitted to the human auditory system so that the vocal system is controlled for the stable vocalization.

Infants have the vocal organs congenitally, however they cannot utter a word. As infants grow they acquire the control methods pertaining to the vocal organs for appropriate vocalization. These get developed in infancy by repetition of trials and errors concerning the hearing and vocalizing of vocal sounds. Any disability or injury to any part of the vocal organs or to the auditory system might cause an impediment in vocalization. People who have congenitally hearing impairments have difficulties in learning vocalization, since they are not able to listen to their own voice.

Auditory impaired patients usually receive a speech training conducted by speech therapists (STs) (Boothroyd, 1988; Boothroyd 1973; Erber et al, 1978; Goldstein & Stark, 1976), however many problems and difficulties are reported. For example, in the training, a patient is not able to observe his own vocal tract, nor the complex articulations of vocal organs in the mouth, and he cannot recognize the validity of his speech articulations nor evaluate the achievement of speech training without hearing the voices. Children regularly take training in a deaf school during a semester, however it is not easy to continue the training during school holidays, and they tend to forget the skills, so that they resume the training again in the beginning of the new semester by repeating the previously-conducted training menus. The most serious problem is that the number of STs is not enough to give speech training to all the patients with auditory impairment. A simple training system or a supporting device that a patient regularly uses for daily speech training by oneself is strongly required.

2. VOCALIZATION TRAINING AND THE RELATED STUDIES

Figure 1 shows two examples of electronic speech training systems, WH-9500 developed by Matsushita Electric Industrial Co., Ltd., and JX-1 developed by Body sonic Co., Ltd. Equipped with a headset with a microphone, WH-9600 directs the difference of sound features together with an estimated vocal tract shape on the display, so that a trainee could understand his own vocalization visually. The system is large and requires technical knowledge and complex settings, and it is difficult for an individual patient to settle it at home. JX-1, on the other hand, consists of two vibration units, one for hands and the other for a body. The trainee inputs a voice to the system via a microphone, and the system presents the phonetic difference between the able-bodied voice and the trainee voice by vibration patterns. It has an advantage of simple operations without special knowledge, however the system is not able to direct how a trainee articulates the vocalization for the better vocalization during the training. By examining the problems of the conventional training systems, the authors are constructing an interactive training system, by which a patient engages in speech training in any occasion, at any place, without special knowledge.

(a) WH-9600 (Matsushita Electric Industrial Co., Ltd.) *(b) JX-1 (Bodysonic Co., Ltd.)*

Figure 1. *Examples of electronic speech training systems.*

The authors are developing a talking robot by reproducing a human vocal system mechanically based on the physical model of human vocal organs (Kitani et al, 2008). The robot consists of motor-controlled vocal organs such as vocal cords, a vocal tract and a nasal cavity to generate a natural voice imitating a human vocalization. For the autonomous acquisition of the robot's vocalization skills, an adaptive learning using an auditory feedback control is employed. In the previous study, the talking robot was applied to the training system of speech articulation for the hearing impaired children, since the robot is able to reproduce their vocalization and to teach them how it is improved to articulate the vocal organs for generating clear speech. The basic concept of the training system is shown in Figure 2.

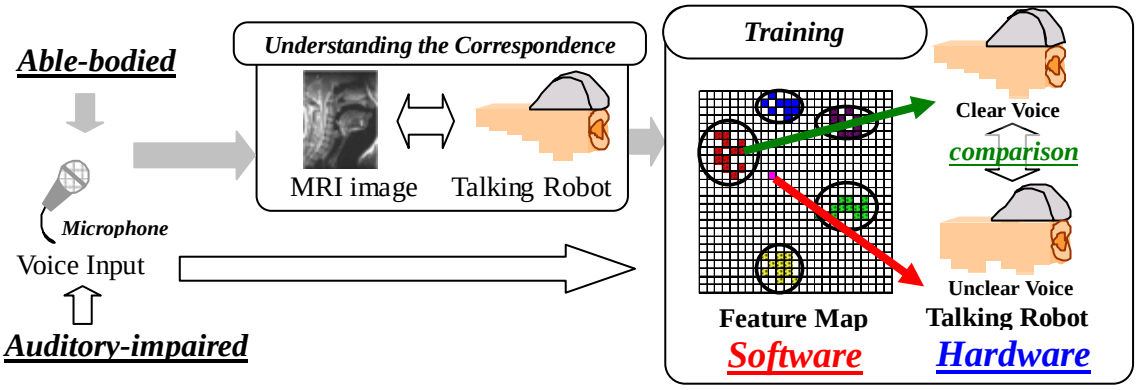


Figure 2. *Scheme of speech training system.*

The training is given by two approaches; one is to employ the talking robot for directing the shape and the motion of the vocal organs (hardware training), and the other is to use a topological map for presenting the difference of phonetic features of trainee's voices (software training). Prior to the experiments, the MRI images are presented to subjects to help understand the correspondence of human inner-mouth shapes with the robotic vocal tract shapes.

Firstly, an ideal vocal tract shape for a clear vocalization is presented to a trainee by the talking robot, and then the trainee tries to mimic the articulation of the vocalization by referring to the robot motion. Simultaneously, by listening to the trainee's voices, the robot reproduces the trainee's vocal tract shapes, and directs how the trainee's voice would be clarified by the change of articulatory motions, by intensively showing the different articulatory points. The trainee compares his own vocal tract shape and the ideal vocal

Proc. 8th Intl Conf. Disability, Virtual Reality & Associated Technologies
Viña del Mar/Valparaíso, Chile, 31 Aug. – 2 Sept. 2010

tract shape, both of which are shown by the articulatory motions of the robot, and tries to reduce the difference of the articulations. At the same time, the system also presents phonetic features using a phonetic topological map, in which the relations of the phonetic characteristics of trainee's voices and the target voices are displayed. By repeating the utterance and listening, the trainee would be able to recognize the similarity of the phonetic features presented as the topological distance between his voice and the target voice, and tries to reduce the distance. In the training, a trainee repeats these training processes for learning 5 vowels.

For assessing the effectiveness of constructed system, we conducted an experiment in the Kagawa prefectural school for the deaf, and significant results were obtained. The system used the 2D feature map for locating the phonetic characteristics, and we found it could not locate the phonetic characteristics properly. Additionally, the neural network which we used for associating phonetic characteristics with motor control commands required a lot of control parameters and reactive settings for the better learning. With these reasons, an algorithm which enables the phonetic characteristics to be located properly in the topological map, and associates the phonetic characteristics with motor controls by reduced control parameters is strongly required. In this study, a new algorithm using two SOMs, called a dual-SOM, is introduced for the autonomous learning. In the following chapters, the construction of the robotic training system and the training experiments are described, together with the performance of the dual-SOM learning for the robotic voice articulations.

3. CONSTRUCTION OF A TALKING ROBOT

The talking robot mainly consists of an air pump, artificial vocal cords, a resonance tube, a nasal cavity, and a microphone connected to a sound analyzer, which, respectively, correspond to a lung, vocal cords, a vocal tract, a nasal cavity, and a human audition, as shown in Figure 3. An air flow from the pump is led to the vocal cords via an airflow control valve, which works for the control of the voice volume. The resonance tube as a vocal tract is attached to the vocal cords for the manipulation of resonance characteristics. The nasal cavity is connected to the resonance tube with a rotary valve settled between them. The sound analyzer plays a role of the auditory system, and realizes the pitch extraction and the analysis of resonance characteristics of generated sounds in real time, which are necessary for the autonomous learning of vocalization skill. The system controller manages the whole system by listening to the vocalized sounds and calculating motor control commands, based on the auditory feedback control mechanism employing a neural network learning. The relation between the phoneme characteristics of generated voice and motor control parameters is stored in the system controller, which is referred to in the generation of speech performance.

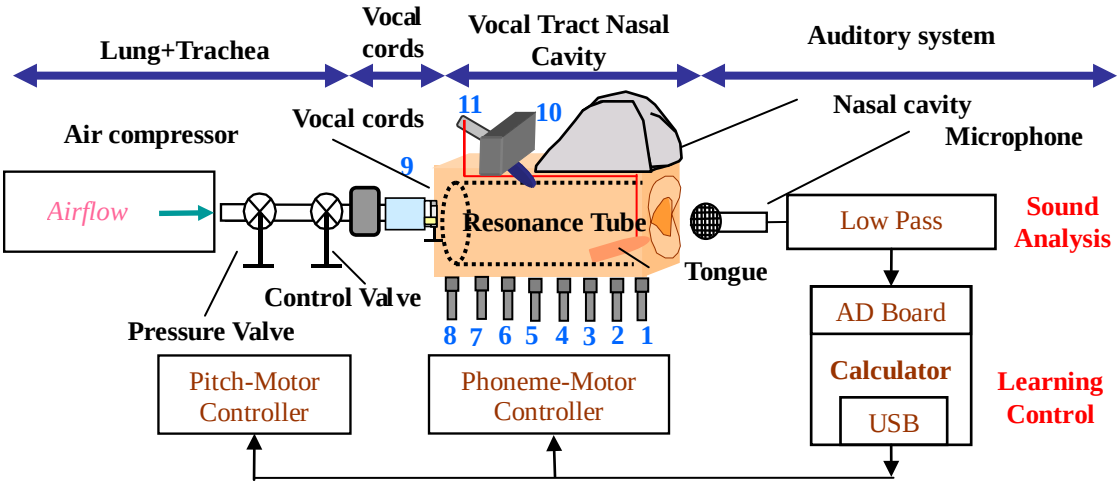


Figure 3. *Construction of a talking robot.*

The characteristics of a glottal wave which determines the pitch and the volume of human voice is governed by the complex behavior of the vocal cords. It is due to the oscillatory mechanism of human organs consisting of the mucous membrane and muscles excited by the airflow from the lung. Although several researches about the computer simulations of the movements are found, we are trying to generate the wave by a mechanical model. We employed an artificial vocal cord used by people who had to remove their vocal cords because of a glottal disease. The vibration of a rubber with the width of 5mm attached over a plastic body makes vocal sound source. The tension of the rubber can be manipulated by applying tensile force. We measured the relationship between the tensile force and the fundamental frequency of a vocal sound generated by the artificial vocal cord (Higashimoto & Sawada, 2003). The fundamental frequency varies

from 110 Hz to 350 Hz by the manipulations of a force applying to the rubber. While, the relation between the produced frequency and the applied force is not stable but tends to change with the repetition of experiments due to the fluid dynamics. The artificial vocal cord is, however, considered to be suitable for our system not only because of its simple structure, but also its frequency characteristics to be easily controlled by the tension of the rubber and the amount of airflow. For the adjustments of fundamental frequency and volume, two motors are employed: one is to manipulate a screw of an airflow control valve, and the other is to apply a tensile force to the rubber of the vocal cord for the tension adjustment.

The human vocal tract is a non-uniform tube about 170 mm long in man. Its cross-sectional area varies from 0 to 20 cm^2 under the control for vocalization. A nasal tract with a total volume of 60 cm^3 is coupled to the vocal tract. Nasal sounds such as /*m*/ and /*n*/ are normally excited by the vocal cords and resonated in the nasal cavity. Nasal sounds are generated by closing the soft palate and lips, not to radiate air from the mouth, but to resonate the sound in the nasal cavity. The closed vocal tract works as a lateral branch resonator and also has effects of resonance characteristics to generate nasal sounds. Based on the difference of articulatory positions of tongue and mouth, the /*m*/ and /*n*/ sounds can be distinguished with each other.

In the mechanical system, a resonance tube as a vocal tract is attached at the sound outlet of the artificial vocal cords. It works as a resonator of a source sound generated by the vocal cords. It is made of a silicone rubber with the length of 180 mm and the diameter of 36 mm, which is equal to 10.2 cm^2 by the cross-sectional area. The silicone rubber is molded with the softness of human skin, which contributes to the quality of the resonance characteristics. In addition, a nasal cavity made of a plaster is attached to the resonance tube to vocalize nasal sounds like /*m*/ and /*n*/.

By actuating displacement forces with stainless bars from the outside, the cross-sectional area of the tube is manipulated so that the resonance characteristics are changed according to the transformations of the inner areas of the resonator. DC motors are placed at 8 positions x_j (j = 1-8) from the intake side of the tube to the outlet side, and the displacement forces P_j (x_j) are applied according to the control commands from the phoneme-motor controller. A nasal cavity is coupled with the resonance tube as a vocal tract to vocalize human-like nasal sounds by the control of mechanical parts. A rotational valve as a role of the soft palate is settled at the connection of the resonance tube and the nasal cavity for the selection of nasal and normal sounds. For the generation of nasal sounds /*n*/ and /*m*/, the rotational valve is open to lead the air into the nasal cavity. By closing the middle position of the vocal tract and then releasing the air to speak vowel sounds, /*n*/ consonant is generated. For the /*m*/ consonants, the outlet part is closed to stop the air first, and then is open to vocalize vowels. The difference in the /*n*/ and /m/ consonant generations is basically the narrowing positions of the vocal tract. In generating plosive sounds such as /*p*/, /*b*/ and /*t*/, the mechanical system closes the rotational valve not to release the air in the nasal cavity. By closing one point of the vocal tract, air provided from the lung is stopped and compressed in the tract. Then the released air generates plosive consonant sounds like /*p*/ and /*t*/. The robot also has a silicone-molded tongue, which is made by referring to the shape and size of a human. A string is attached to the tongue, and at the other end of the string, a servo motor is connected for the manipulation of the up-down motion, to articulate the vocalization of /*l*/ sounds.

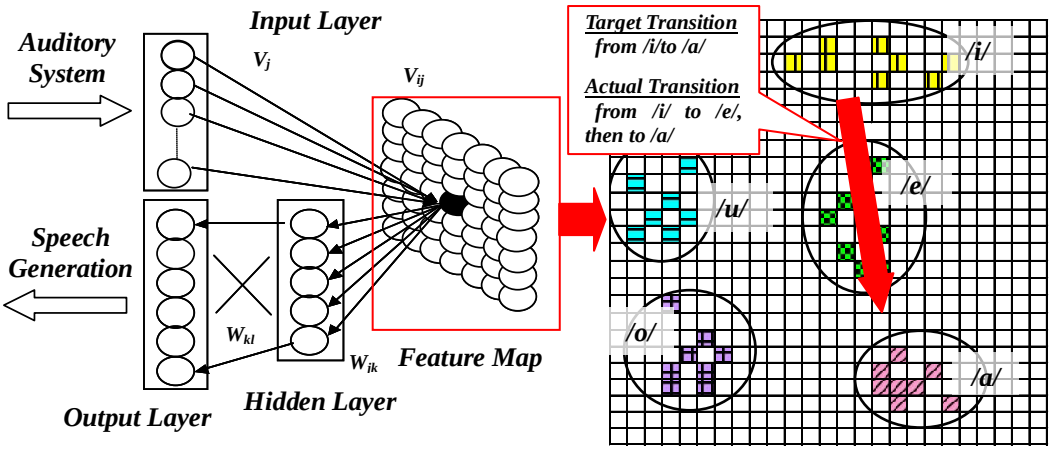
4. METHOD OF AUTONOMOUS VOICE ACQUISITION

In the previous work, the Self-organizing Neural Network (SONN) was employed for associating phonetic characteristics with motor control commands (Sawada & Nakamura, 2004; Sawada, 2007). The SONN has a 2 dimensional feature map for locating phonetic characteristics as shown in Figure 4. In each cell on the feature map, the phonetic characteristics and motor control commands are buried by establishing topological relations, and with neighborhood learning of the SOM, similar phonetic characteristics are located close with each other.

As shown in a bold arrow in the figure 4 (b), when two cells are chosen from one cell to the other on the feature map (in this case, the voice from /*i*/ to /*a*/), the voice transition is obtained by selecting cells under the path connecting the two cells. However, if another vocal sound is situated on the path (voice area /*e*/ is situated in the map), the different voice is vocalized during the transition. This problem possibly occurs in the improper learning of the phonetic characteristics, and for solving this problem a new algorithm which enables the voice learning to locate all the voice features in proper locations on the map is required.

A 3D SOM which has 3 dimensional mapping space is introduce in this study for properly locating the phonetic characteristics. By applying 3 dimensional space, the characteristics are located 3 dimensionally, and the probability that the SOM generates improper locations would be decreased. Figure 5 shows a

mapping result of the phonetic characteristics. The phonetic parameters are well mapped three dimensionally, and five vowels are categorized with one another by the learning of the SOM.



a) *Structure of Self-organizing Neural Network* *b*) *Feature Map*

Figure 4. *Previous learning method.*

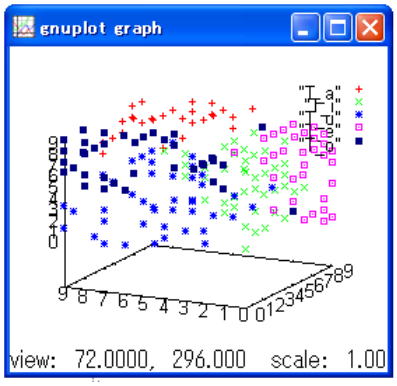


Figure 5. *Mapping result.*

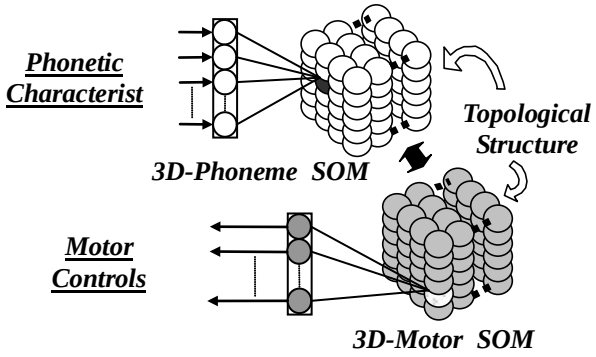


Figure 6. *Structure of dual-SOM.*

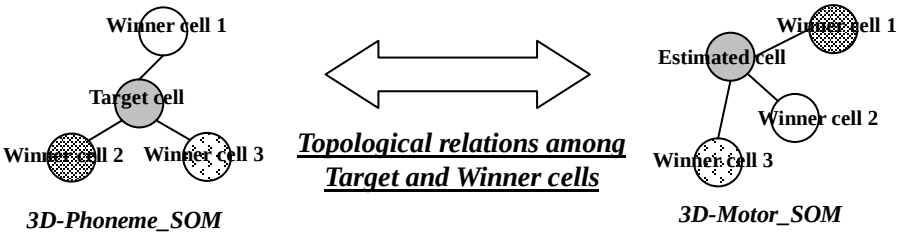


Figure 7. *Association of 3D-Phoneme_SOM with 3D-Motor_SOM.*

A dual-SOM is employed to associate the motor control commands of the robot with the phonetic characteristics of generated voices. The structure of the dual-SOM is shown in Figure 6, which consists of two self-organizing maps. One is a 3D-Motor_SOM, which describes the topological relations of various shapes of the vocal tracts, in which close shapes are arranged in close locations with each other, and the other is 3D-Phoneme_SOM, which learns the relations among phonetic features of generated voices. The talking robot generates various voices by changing its own vocal tract shapes. Generated voices and vocal tract shapes have the physical correspondence, since different voices are produced by the resonance phenomenon of the articulated vocal tract. This means that similar phonetic characteristics are generated by similar vocal tract shapes. By adaptively associating the 3D-Phoneme_SOM with the corresponding 3D-Motor_SOM, we could expect that the talking robot autonomously learns the vocalization by articulating its vocal tract.

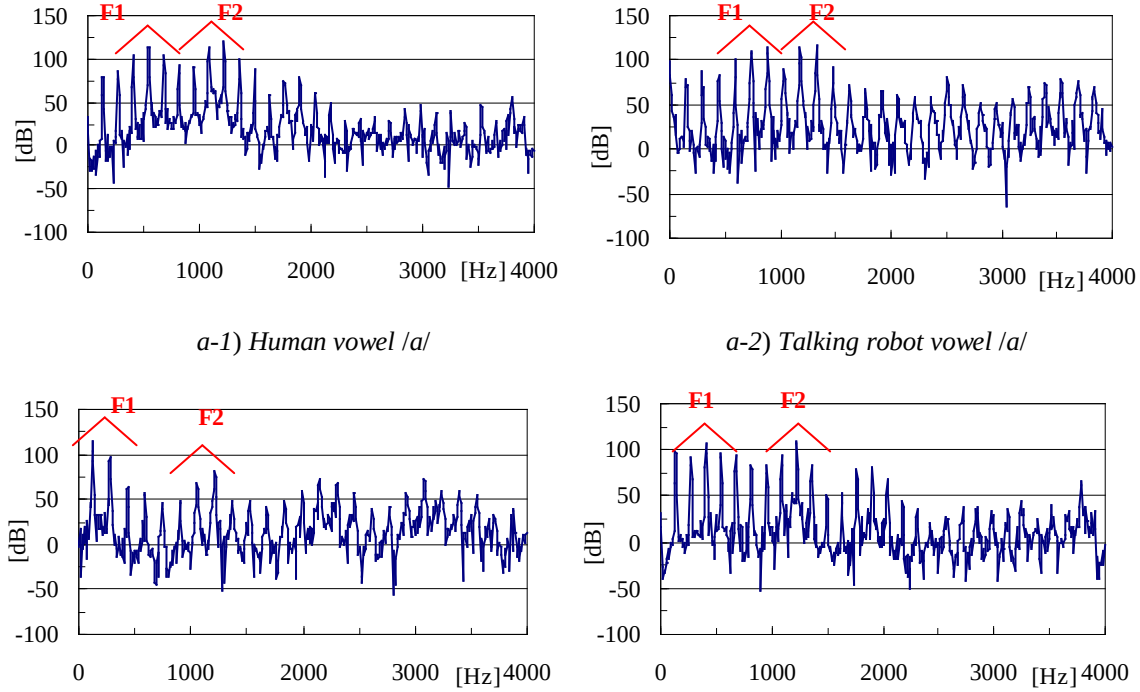
In the learning phase, the motor control commands and the corresponding phonetic characteristics consisting of 9th order LPC cepstra are obtained by random articulations of the talking robot, and are

inputted to the 3D-Motor_SOM and the 3D-Phoneme_SOM, respectively. The topological structures are autonomously established by the neighborhood learning on each SOM, so that similar patterns are located close with each other, and different patterns are located apart. The differences among patterns appear as the norm information in the three dimensional space in the SOM, so we tried to associate the two maps with each other by referring to the norms among a target cell and winner cells, which are shown in Figure 7. First, in the 3D-Phoneme_SOM, the distances from a target cell to the selected 3 winner cells are calculated, and the topological relations among 4 cells are also obtained. Then, by applying the topological relations to the 3D-Motor_SOM, the location of a cell from the corresponding 3 winner cells is estimated. The estimated location in the 3D-Motor_SOM would generate the corresponding vocal tract shape given by the phonetic features of the inputted sound.

5. LEARNING RESULT OF VOICE ARTICULATION

For the validation of the voice learning, we conducted an experiment of the robotic voice generation. Figure 8 shows the results of acquired spectra of /*a*/ and /*u*/ vowels, in comparison with actual human voices, and we found that the phonetic characteristics of Japanese vowels were well reproduced. Human vowel /*a*/ has the first formant in the frequency range from 500 to 900 Hz and the second formant from 900 to 1500 Hz, and the robotic voice also presents the two formants in the range. In the listening experiments, most of the subjects pointed out that the generated voices have similar phonetic characteristics to the human voices. These results show that the vocal tract made by silicone rubber has the tolerance of generating human-like vocalization, and application of the dual-SOM for the voice acquisition was well achieved.

Figure 9 shows the results of the transition of Japanese vowels from */u/* to */o/*. The figure *a*) shows the transition of phonetic characteristics buried in the 3D-Phoneme_SOM, and the figure *b*) shows the transition of motor control parameters obtained by the 3D-Motor_SOM. The abscissas show the time steps of the transition, and the ordinates show the phonetic characteristics and the motor control values, respectively. As the phonetic characteristics changed its values from */u/* to */o/*, the corresponding motor control values changed gradually as expected. The table *c*) shows the selected cells on the 3D-Phoneme_SOM and 3D-Motor_SOM through the transition from */u/* to */o/*. The left side column shows the time steps of the transition, and X Y Z shows the coordinates of the cells in two SOMs. Through the transition, selected cells were gradually changing its positions from */u/* to */o/*. This result confirms the cells on two SOMs were properly chosen. By the use of the three dimensional space for the mapping, the transitions were properly generated, and the results presented that the associations between two SOMs were well achieved.



a-1) *Human vowel /a/*

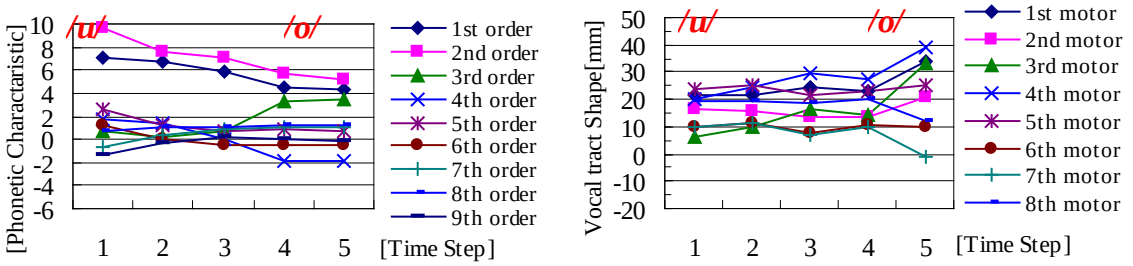
a-2) *Talking robot vowel /a/*

b-1) *Human vowel /u/*

b-2) *Talking robot vowel /u/*

Figure 8. *Comparison of spectra.*

Proc. 8th Intl Conf. Disability, Virtual Reality & Associated Technologies
Viña del Mar/Valparaíso, Chile, 31 Aug. – 2 Sept. 2010



a) 3D-Phoneme_SOM *b) 3D-Motor_SOM*

	3D-Phoneme_SOM			3D-Motor_SOM		
	X	Y	Z	X	Y	Z
1	2	7	9	4	4	4
2	2	6	8	3	4	2
3	2	5	7	5	5	3
4	2	4	6	4	4	3
5	2	3	5	4	6	5

c) Selected cells on two SOMs

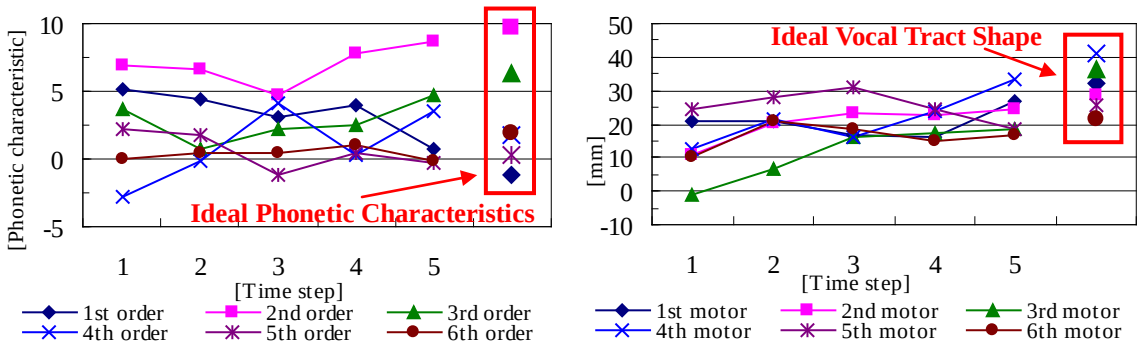
Figure 9. *Transition from /u/ to /o/ voices.*

6. VOCALIZATION TRAINING BY APPLYING Dual-SOM

A training experiment was conducted in Kagawa Prefectural School for the Deaf. The training consists of two parts, one is to employ the robot for directing the articulatory motion of the vocal organs for a vocalization (hardware training), and the other is to present a 3D feature map on a computer display for presenting the difference of phonetic features of trainee's voices (software training). The trainee compares his own vocal tract shape with the ideal vocal tract shapes, both of which are shown by the articulatory motions of the robot, and tries to reduce the difference of the articulations. The system also presents phonetic features using the 3D topological map, in which the trainee's voice and the target voices are visually referred to by the topological relations. During the repetition of uttering and listening, the trainee recognizes the topological distance between his voice and the target voice, and tries to reduce the distance. In the training, a trainee repeats these training processes for learning 5 vowels. Six high school students and four junior-high school students (10 students in total) were engaged in the experiment.

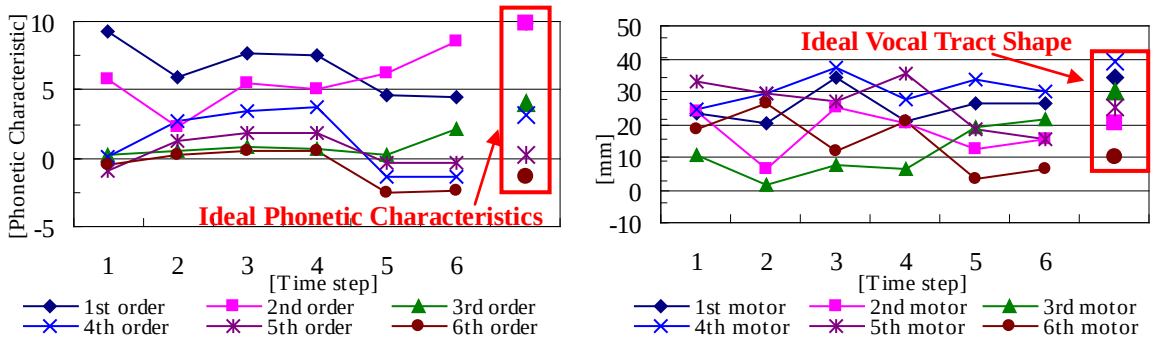
Figure 10 shows the results of training /i/ and /o/ vowels for subject 1. In the figures of *a*) and *b*), the abscissas show the time steps of the training, and the ordinates show the phonetic characteristics and the vocal tract shapes reproduced by the talking robot, respectively. As the training proceeds, the phonetic characteristics and the corresponding vocal tract shapes of the vowels were getting close to the ideal ones given by able-bodied speech. These results confirmed that the subject 1 learned the vocalization properly. The figure *c*) shows the training results of the 5 vowels. In the 3D Feature Map shown in figure *c*), each marker presents the locations of 5 vowels, where same colors show the same vowels. The greater markers present the averaged locations of 5 vowels given by able-bodied speech, and small markers present the vocalization of subject 1. Table *d*) shows the selected cells in the 3D feature map, in which the left side column shows the 5 vowels, and X Y Z shows the coordinates of the cells in the map. Before the training, the locations of the selected cells were far from the able-bodies speech, however after the training, selected cells were located closer to the coordinates of the ideal ones. The results verified that the subject 1 successfully acquired the proper vocalizations by the interactive training.

Figure 11 shows the result of the training of /u/ vowel of subject 2. The abscissas and the ordinates in the figures *a*) show the same parameters as presented in the figure 10. Both phonetic features and reproduced vocal tract shapes given by the subject 2 voices were far different from the ideal values, and selected cells on the 3D feature map were located far from the vowel /u/ ideal location. The subject claimed the difficulty of understanding the relations of phonetic features presented three dimensionally, and he pointed out the necessity of much intuitive understanding of articulations in the mouth.



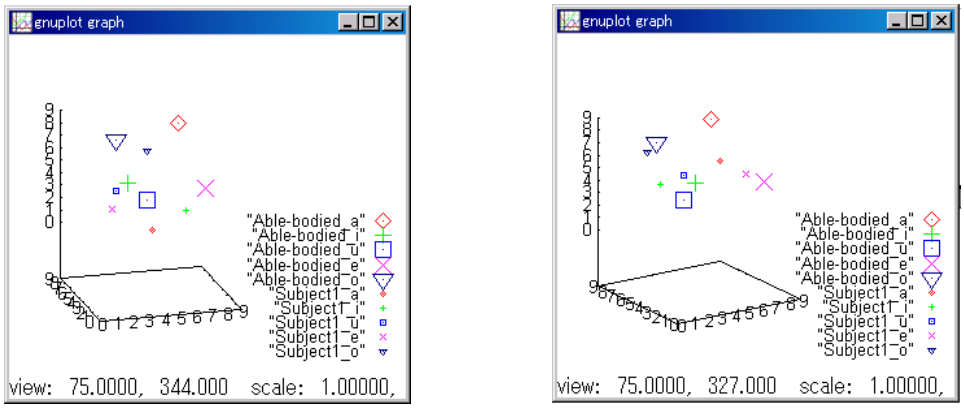
a-1) Phonetic characteristics *a-2) Vocal tract shapes*

a) Characteristics of training /i/ on 3D feature map



b-1) Phonetic characteristics *b-2) Vocal tract shape*

b) Characteristics of training of /o/ on 3D feature map



c-1) Before Training *c-2) After Training*

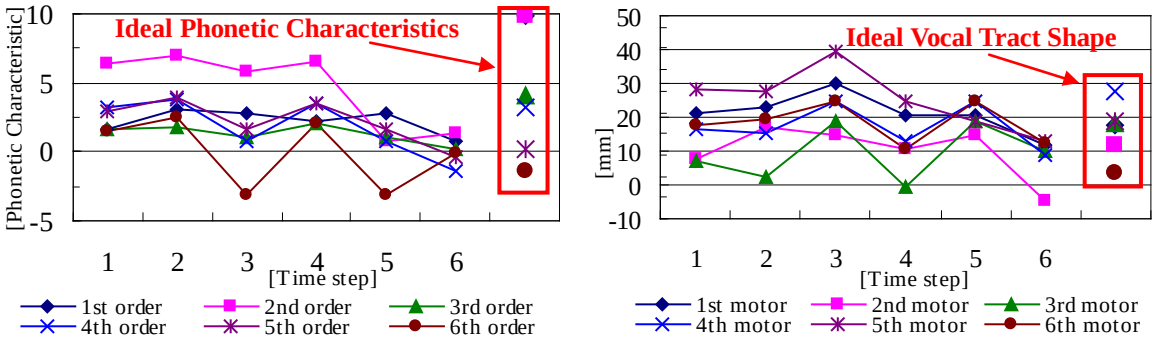
c) 3D feature map of Subject 1

Vowels	Before Training			After Training			Able-bodied Vocalization		
	X	Y	Z	X	Y	Z	X	Y	Z
/a/	5	6	0	7	6	5	7	7	8
/i/	6	2	3	0	2	6	2	1	6
/u/	3	7	3	5	7	4	5	7	2
/e/	1	1	4	7	3	5	7	1	5
/o/	5	7	6	3	8	6	3	7	7

d) Selected cells on 3D Feature Map

Figure 10. *Training Result of Subject 1.*

Proc. 8th Intl Conf. Disability, Virtual Reality & Associated Technologies
Viña del Mar/Valparaíso, Chile, 31 Aug. – 2 Sept. 2010



a-1) Phonetic characteristics *a-2) Vocal tract shape*

a) Characteristics of training of /u/ on 3D feature map

Vowels	Before Training			After Training			Able-bodied Vocalization		
	X	Y	Z	X	Y	Z	X	Y	Z
/a/	1	9	1	7	6	8	7	7	8
/i/	2	4	9	2	2	5	2	1	6
/u/	5	8	8	7	9	8	5	7	2
/e/	4	0	0	2	2	4	7	1	5
/o/	0	9	0	4	4	8	3	7	7

b) Selected cells on 3D Feature Map

Figure 11. *Training Result of Subject 2*

By the robotic speech training, 7 students out of 10 could successfully improve their vocalization, and the left 3 students partly learned the articulations for the better vocalization. These results verified that the speech training employing the robot and the intuitive directions helped the trainees understand how to articulate the speech for the clear vocalization. In the questionnaire after the training, most subjects answered that the training with the robot was fun, and all of them gave the positive participation to the training.

7. CONCLUSION

In the paper a new learning method of the autonomous vocalization for a talking robot employing a dual-SOM was introduced, and the robot was applied to the interactive training for the auditory-impaired people. By analyzing the problem of previously-introduced algorithm using a SONN, we constructed a new algorithm using two SOMs to establish topological relations of the phonetic characteristics with motor-control parameters in a three dimensional space. By utilizing the autonomous learning of phonetic features of vocal sounds, the robot was employed in the speech training for auditory-impaired students. The experimental results verified the effectiveness of the interactive training using a robot and the intuitive directions of vocal characteristics. We are now working to construct the better training system which could estimate and show the detailed phonetic characteristics of the trainee.

Acknowledgements: This work was partly supported by the Grants-in-Aid for Scientific Research, the Japan Society for the Promotion of Science (No. 21500517). The authors would like to thank Dr. Yoichi Nakatsuka, the director of the Kagawa Prefectural Rehabilitation center for the Physically Handicapped, Mr. Tomoyoshi Noda, the speech therapist and teacher of Kagawa Prefectural School for the Deaf, and the students of the school for their helpful supports for the experiment and the useful advice.

8. REFERENCES

A. Boothroyd: “Hearing Impairments in Young Children”, A. G. Bell Association for the Deaf, 1988.

A. Boothroyd: “Some experiments on the control of voice in the profoundly deaf using a pitch extractor and storage oscilloscope display”, IEEE Transactions on Audio and Electroacoustics, Vol.21, No.3, pp. 274-278, 1973.

N. P. Erber and C. L. de Filippo: “Voice/mouth synthesis and tactual/visual perception of /pa, ba, ma/”, Journal of the Acoustical Society of America, Vol.64, No.4, pp.1015-1019, 1978.

M. H. Goldstein and R. E. Stark: “Modification of vocalizations of preschool deaf children by vibrotactile and visual displays”, Journal of the Acoustical Society of America, Vol.59, No.6, pp.1477-81, 1976.

M. Kitani, Y. Hayashi and H. Sawada: “Interactive training of speech articulation for hearing impaired using a talking robot”, International Conference on Disability, Virtual Reality and Associated Technologies, pp.293-301, 2008.

T. Higashimoto and H. Sawada: “A Mechanical Voice System: Construction of Vocal Cords and its Pitch Control”, International Conference on Intelligent Technologies, pp. 762-768, 2003.

H. Sawada and M. Nakamura: “Mechanical Voice System and its Singing Performance”, IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1920-1925, 2004.

H. Sawada: “Talking Robot and the Autonomous Acquisition of Vocalization and Singing Skill”, Chapter 22 in Robust Speech Recognition and Understanding, Edited by Grimm and Kroschel, pp.385-404, June 2007.