

Powered Wheelchair Control Using Acoustic-Based Recognition of Head Gesture Accompanying Speech

Akira Sasou¹

¹National Institute of Advanced Industrial Science and Technology (AIST)

a-sasou@aist.go.jp

Abstract

In this paper, we propose the novel interface for powered wheelchair control using the acoustic-based recognition of head gesture accompanying speech. A microphone array mounted on a wheelchair localizes the position of the user's voice. Because the localized position of the user's voice almost corresponds with that of the mouth, the tracking of the head movements accompanying speech can be achieved by means of the microphone array. The proposed interface does not require disabled people to wear any microphones or utter recognizable voice commands, but requires only two capabilities: the ability to move the head and the ability to utter an arbitrary sound. In our preliminary experiments, five subjects performed six kinds of head gestures accompanying speech. The head gestures of each subject were recognized using the models trained from the other subjects' data. The average recognition accuracy was 99.7 %.

Index Terms: head gesture recognition, microphone array

1. Introduction

Although various voice-driven wheelchairs have already been developed to enable disabled people to move independently, conventional voice-driven wheelchairs had some associated problems [1-4]. Conventional voice-driven wheelchairs employed a headset microphone that can record the user's voice command in a higher Signal-to-Noise Ratio (SNR), even in the presence of surrounding noise, and can achieve sufficient speech recognition accuracy. However, users need to put on this headset microphone each time they use the wheelchair. In addition, when the headset microphone moves away from the position of the mouth, users need to be able to adjust the position of the headset microphone by themselves. These actions are not always easy, especially for the hand disabled, who are one of the major users of this wheelchair. Since such users need noncontact and nonconstraining interfaces for controlling the wheelchair, we appraised headset microphones as impractical. Conversely, it is also well known that the speech recognition accuracy drastically degrades when the microphone is placed far from the user because surrounding noises can easily interfere with the user's voice.

In order to overcome such difficulties, we developed a noise robust speech recognition system for a voice-driven wheelchair in our previous study [5]. To eliminate the need for the user to wear a microphone, we developed a microphone array system that is mounted on a wheelchair as shown in Fig.1. The developed microphone array system can easily distinguish the user's utterances from other voices by localizing the sound source positions instead of using a speaker identification technique. We also adopt a feature compensation technique following to the microphone array system. As a result of combining these two methods, the weak point of the microphone array, which is processing omnidirectional noises, can be compensated for by the feature

compensation method. Consequently, our system can be applied to a variety of noise environments.

On the other hand, there exist some disabled people who can hardly utter voice commands clearly. Such inarticulate commands cause inaccurate speech recognition even if the interfering noises are suppressed. So the voice-driven wheelchair developed in our previous study is not suitable for this kind of disabled people. Some researchers developed gaze or head gesture interfaces for powered wheelchair control based on visual information [6-8]. With the visual-based system, however, the user cannot move his/her head except for controlling the wheelchair.

In this paper, we propose the novel interface for powered wheelchair control using the acoustic-based recognition of head gesture accompanying speech. The microphone array, which was originally developed for the purpose of achieving noise robust speech recognition, needs to localize the position of the user's voice and the arrival directions of surrounding noises to enhance the user's voice by beam-forming. Because the localized position of the user's voice almost corresponds with that of the mouth, the tracking of the head movements accompanying speech can be achieved by means of the microphone array. The proposed interface does not require disabled people to wear any microphones or utter recognizable voice commands, but requires only two capabilities: the ability to move the head and the ability to utter an arbitrary sound. The user can move his/her head freely if they don't say anything, but the user also moves his/her head a little when talking to others. These head gestures during dialogue, however, are distinguishable from those for wheelchair controls. We therefore train a model of the head gestures accompanying dialogue so that the head gestures are allowed during speech.



Figure 1: A wheelchair with the microphone array.

2. User Utterance Localization

Figure 1 shows the wheelchair mounting microphone array we have developed, which consists of two circuit boards. Each circuit board is $W130 \times D10 \times H5$ mm in size and has four omnidirectional silicon microphones soldered in a line at intervals of 3 cm in order to avoid spatial aliasing at frequencies up to 4 kHz. The circuit boards are placed in a diagonal direction on black square sponges on the armrests as shown in Fig.1. Because these black sponges are placed on the edges of the arm rests, the user's head never touches the microphone array system, even during involuntary movements. The sampling frequency is set to 8 kHz. In order to distinguish the user's voice and the surrounding noises, we define the User Utterance Area (UUA) as shown in Fig.2, which has a space of $W20 \times D15$ cm² in front of microphone array. If the sound is localized in the UUA, the system accepts the sound as the user's utterance. When the sound occurs outside of the UUA, the system rejects the sounds as a noise. Therefore, the microphone array can easily distinguish the user's voice from other's voices and/or noises without training procedures such as speaker identification.

We have adopted the MUSIC method [9] for the User Utterance Localization (UUL). We assume that a sound occurring in the UUA is received as a spherical wave by the microphone array. The steering vectors are defined as follows.

$$\mathbf{P}_q = [Px_q, Py_q, Pz_q]^T, q = 1, \dots, 8$$

$$R_q = |\mathbf{P}_q - \mathbf{P}_0| = \sqrt{(Px_q - Px_0)^2 + (Py_q - Py_0)^2 + (Pz_q - Pz_0)^2}$$

$$\tau_q = R_q / v, g_q = g(\omega, R_q) \quad (1)$$

$$\mathbf{a}(\omega, \mathbf{P}_0) = [g_1 e^{-j\omega\tau_1}, \dots, g_8 e^{-j\omega\tau_8}]^T$$

where \mathbf{P}_0 represents the position of the sound source in the UUA, $\mathbf{P}_1 \dots \mathbf{P}_8$ represent the positions of microphones, R_q is the distance between the q th microphone and the sound source and v is sound velocity.

The spatial correlation matrix is defined as

$$\mathbf{R}(\omega) = (1/N) \sum_{n=1}^N \mathbf{y}(\omega, n) \mathbf{y}^H(\omega, n) \quad (2)$$

where $\mathbf{y}(\omega, n) = [Y_1(\omega, n), \dots, Y_8(\omega, n)]^T$, and $Y_q(\omega, n)$ represents the FFT of the n th frame received by the q th microphone. The eigenvalue decomposition of $\mathbf{R}(\omega)$ is given by

$$\mathbf{R}(\omega) = \mathbf{E}(\omega) \mathbf{L}(\omega) \mathbf{E}^{-1}(\omega) \quad (3)$$

where $\mathbf{E}(\omega)$ denotes the eigenvector matrix, which consists of the eigenvectors of $\mathbf{R}(\omega)$ as $\mathbf{E}(\omega) = [\mathbf{e}_1(\omega), \dots, \mathbf{e}_8(\omega)]$, and $\mathbf{L}(\omega)$ denotes the eigenvalue matrix, which is defined as

$$\mathbf{L}(\omega) = \text{diag}(\lambda_1(\omega), \dots, \lambda_8(\omega)) \quad (4)$$

$$\lambda_1(\omega) \geq \dots \geq \lambda_8(\omega)$$

The number of sound sources is estimated from the eigenvalues as follows. First, we evaluate the threshold value, which is defined as

$$T_{egn}(\omega) = \lambda_1^{C_{egn}}(\omega) \times \lambda_8^{(1-C_{egn})}(\omega), 0 < C_{egn} < 1 \quad (5)$$

where C_{egn} is a constant that is adjusted experimentally. The number of sound sources $N_{snd}(\omega)$ is then estimated as the number of eigenvalues larger than the threshold value.

$$\lambda_1(\omega), \dots, \lambda_{N_{snd}}(\omega) \geq T_{egn}(\omega) \quad (6)$$

The eigenvectors corresponding to these eigenvalues form the basis of signal subspace $\mathbf{E}_s(\omega) = [\mathbf{e}_1(\omega), \dots, \mathbf{e}_{N_{snd}}(\omega)]$. The

remaining eigenvectors $\mathbf{E}_n(\omega) = [\mathbf{e}_{N_{snd}+1}(\omega), \dots, \mathbf{e}_8(\omega)]$ are the basis of the noise subspace. User utterances are detected according to the following method. First, we search for the position \mathbf{P}_0 that absolutely maximises the following value in the UUA.

$$Q(\mathbf{P}) = 1 / \sum_{\omega} |\mathbf{a}^H(\omega, \mathbf{P}) \mathbf{E}_n(\omega)|^2, \mathbf{P}_0 = \arg \max_{\mathbf{P} \in \text{UUA}} Q(\mathbf{P}) \quad (7)$$

If the absolute maximum value $Q(\mathbf{P}_0)$ exceeds the threshold value T_{usr} , we judge that the user made a sound. In the actual implementation, the user utterance position is evaluated on a grid of 21×16 in the UUA. Thus, the utterance position is localized with an accuracy of 1 cm along each axis.

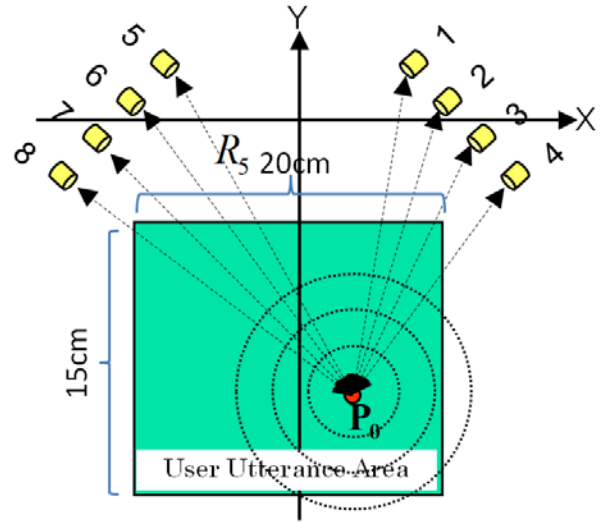


Figure 2: Schematic diagram of wave propagation.

3. Recognition of Head Gesture

In order to achieve recognition of head gesture, we propose a simple and effective method for extracting a feature from the trajectory of X-Y coordinates of the continuous user utterance positions localized by the microphone array. Hereinafter we call this as utterance trajectory. The feature is obtained by calculating the Higher order Local Cross Correlation (HLCC) from the utterance trajectory.

0 th order	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	1 st order	$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$
$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$
$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$
2 nd order	$\begin{bmatrix} 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix}$
$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$
$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$
$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix}$
$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 2 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$
$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$
$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \end{bmatrix}$
$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 1 \end{bmatrix}$			

Figure 3: Example of local patterns.

3.1. Feature Extraction Based on HLCC

In the following, let $x(t)$ and $y(t)$, $t=0, \dots, T-1$, denote the utterance trajectory. First, the bias elements of the utterance trajectory are eliminated according to the following equations.

$$x'(t) = x(t) - \bar{x}, \bar{x} = \sum_{t=0}^{T-1} x(t) / T \quad (8)$$

$$y'(t) = y(t) - \bar{y}, \bar{y} = \sum_{t=0}^{T-1} y(t) / T$$

Second, the time length of the utterance trajectory is normalized to Tn .

$$\begin{aligned} x''(\tilde{t}) &= \{x'(t_b + 1) - x'(t_b)\} \times \Delta t + x'(t_b) \\ y''(\tilde{t}) &= \{y'(t_b + 1) - y'(t_b)\} \times \Delta t + y'(t_b) \end{aligned} \quad (9)$$

$$t_a = T \cdot \tilde{t} / T_n, t_b = \lfloor t_a \rfloor, \Delta t = t_a - t_b$$

The HLCC is then calculated according to

$$\begin{aligned} f(\mathbf{m}) &= \sum_{\tilde{t}=0}^{T_n-1} \{x''(\tilde{t})\}^{p_0} \{x''(\tilde{t}-1)\}^{p_1} \dots \{x''(\tilde{t}-K)\}^{p_K} \\ &\quad \times \{y''(\tilde{t})\}^{q_0} \{y''(\tilde{t}-1)\}^{q_1} \dots \{y''(\tilde{t}-K)\}^{q_K} \end{aligned} \quad (10)$$

where the matrix \mathbf{m} represents the local pattern, given by

$$\mathbf{m} = \begin{bmatrix} p_K & p_{K-1} & \dots & p_0 \\ q_K & q_{K-1} & \dots & q_0 \end{bmatrix} \quad (11)$$

In the above equations, $(K+1)$ represents the length of the local pattern, and the order of the local pattern is defined as $\sum_{k=0}^K (p_k + q_k) - 1$. The HLCC feature vector then is constructed from the HLCCs calculated by using all the local patterns as

$$\mathbf{f} = [f(\mathbf{m}_1) \quad f(\mathbf{m}_2) \quad \dots \quad f(\mathbf{m}_D)]^T \quad (12)$$

where D is the number of the local patterns.

Figure 3 shows the local patterns of the length 3, and the orders in a range from 0^{th} to 2^{nd} , where forty nine patterns exist.

3.2. Subspace Method Based Recognition

The training procedure consists of the following steps. First we collect a number of utterance trajectories of each head gesture. Let $\mathbf{f}_{g,i}$ represent the HLCC feature vector calculated from the i th utterance trajectory of the g th head gesture, and let N_g represent the number of the utterance trajectories of the g th head gesture. Second, the principal component vectors are obtained by the principal component analysis in order to construct the subspace of each head gesture as follows. The auto-correlation matrix is given by

$$\mathbf{A}_g = \sum_{i=1}^{N_g} \mathbf{f}_{g,i} \cdot \mathbf{f}_{g,i}^T / N_g \quad (13)$$

Let $\lambda_{g,k}$ and $\mathbf{u}_{g,k}$, $k=1, \dots, D$, represent the eigen values and the corresponding eigen vectors of the auto-correlation matrix respectively, where the eigen values are sorted in descending order. By means of the cumulative proportion given by

$$\eta_{g,K} = \sum_{k=1}^K \lambda_{g,k} / \sum_{k=1}^D \lambda_{g,k} \quad (14),$$

the dimension of the each subspace is determined according to

$$K_g = \min \{K | \eta_{g,k} \geq Q\} \quad (15).$$

With the orthonormal bases $\mathbf{U}_g = [\mathbf{u}_{g,1} \dots \mathbf{u}_{g,K_g}]$ of the subspace, the projection matrix is given by

$$\mathbf{P}_g = \mathbf{U}_g \cdot \mathbf{U}_g^T \quad (16).$$

In the recognition procedure, we first evaluate the HLCC feature vector \mathbf{f} from the utterance trajectory. The squared-norm of the projection to each subspace is then evaluated as

$$l_g = \|\mathbf{P}_g \cdot \mathbf{f}\|^2 = \|\mathbf{U}_g^T \cdot \mathbf{f}\|^2 \quad (17).$$

We adopt, as the recognition result, the \hat{g} th head gesture that maximizes Eqn. (17) as

$$\hat{g} = \arg \max_g l_g \quad (18).$$

4. Experimental Results

In the following, we first present the evaluation results of the UUL accuracy by means of the microphone array, and then present the recognition accuracy of head gestures based on the HLCC features.

4.1. Accuracy of User Utterance Localization

First, we evaluated the user utterance localization accuracy of the developed system by comparing the results with those of a magnetic 3D positioning sensor. The clean speech signals were recorded with five subjects sitting on the wheelchair and making utterances in a silent room. The subjects were able-bodied, because the purpose of the experiments was to assess the accuracy of user utterance localization. Each subject made utterances around twenty-six times at arbitrary positions in the UUA. In the localization process, the space of the UUA was set to $20 \times 15 \text{ cm}^2$ and the user utterance localization was executed on a $1 \times 1 \text{ cm}$ grid.

We evaluated the error distances on the X and Y axes between the positions estimated by the proposed system and magnetic 3D positioning sensor. Table 1 shows the averages and standard deviations of the error distances on the X and Y axes, which show that the Y axis positioning accuracy of the proposed system tends to degrade more than that on the X axis. This is because the phase changes due to user utterance position changes are smaller along the Y axis than along the X axis.

Table 1. Evaluation Results of UUL accuracy.

	Avg. [cm]	Std. [cm]
X axis	-0.02	1.40
Y axis	0.16	2.01

Next, we evaluated the noise robustness of the user utterance localization. We used clean speech signals and environmental noises separately, and then mixed their digital signals together at eight different SNR levels (20 dB, 15 dB, 10 dB, 5 dB, 0 dB, -5 dB, -10 dB, -15 dB), using a computer to generate noise-corrupted speech signals. The clean speech signals are the same as those in the above experiment. The environmental noises were recorded by moving the wheelchair around in 15 places: 1. near a kindergarten, 2. a construction site near a train, 3. under train rails, 4. in front of an amusement arcade, 5. a restaurant, 6. a building under construction, 7. a public office, 8. in a windy location, 9. along a big street, 10. a road crossing, 11. in front of a drug store, 12. a construction site, 13. a shop, 14. in front of a station and 15. in front of a ticket gate.

Figure 4 shows evaluation results of utterance detection in noisy environments. The insertion errors mean the system detected utterances by mistake, while deletion errors mean the system could not detect utterances. In this evaluation, we need to know how much the noise interference degrades the user utterance localization accuracy. We thus evaluated the error distances along the X and Y axes between the positions estimated from the clean speech signal and noise-corrupted speech signal by the proposed system. Figure 5 represents the error distances on the X and Y axes.

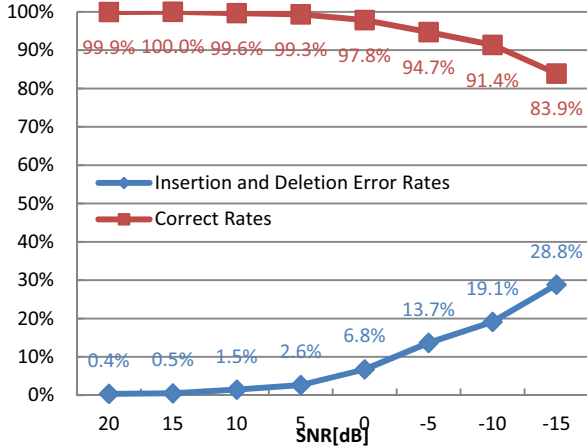


Figure 4: Evaluation results of utterance detection in noisy environments.

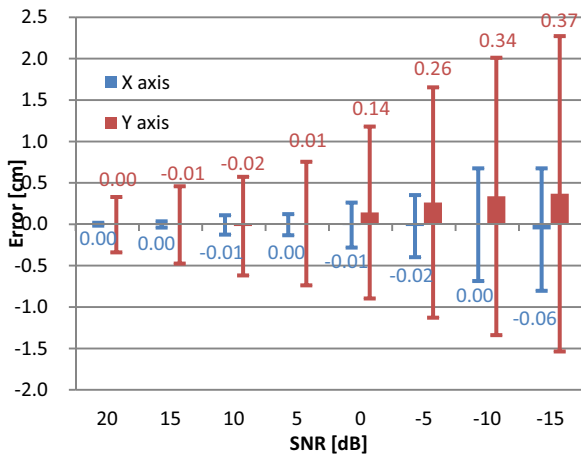


Figure 5: User utterance localization accuracy in noisy environments.

4.2. Accuracy of Head Gesture Recognition

In order to evaluate the recognition accuracy, we collected six kinds of head gestures from five subjects. The subjects made a fricative sound while moving their heads. The head gestures collected in this experiment are as follows. Each subject tilts the head to the forward and to the backward, and faces to the right and to the left. These four head gestures mean to move the wheelchair in each direction. In addition, each subject shakes the head twice between the right and the left, and shakes the head twice between the forward and the backward. These two head gestures mean to stop the wheelchair. The subjects made the same head gestures twenty times.

The HLCC feature vectors were obtained by using the local patterns of the lengths ranging from 6 to 16. The maximum order was set to 2, which means that the local patterns include those of the orders in a range from 0th to 2nd. The T_n in Eqn. (9) was set to 30, 40 and 50, and the Q in Eqn. (15) was set to 0.99999. The recognitions were conducted in an open test manner, that is, the training procedure was conducted based on the data of four subjects, and based on this the remaining subject was tested. Figure 6 shows the average recognition accuracy across the five subjects. In the case where the length of the local pattern was 12 and the T_n in Eqn. (9) was 30, the best score of 99.7 % was obtained.

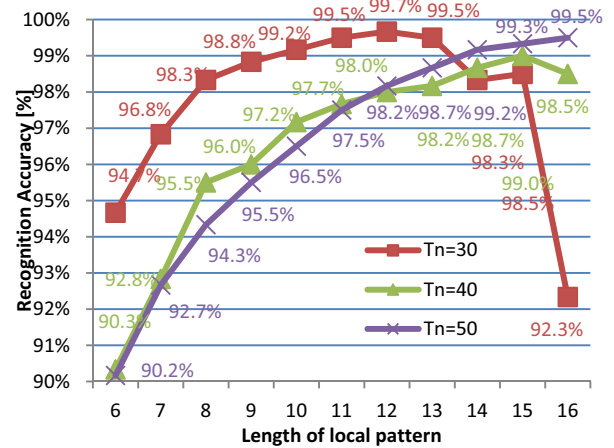


Figure 6: Recognition accuracy of the head gestures.

5. Conclusions

We have proposed the novel interface for controlling a powered wheelchair. The proposed interface adopts the microphone array, which is mounted on the wheelchair, in order to localize the user utterance position. The utterance trajectory obtained from the head gesture accompanying speech is recognized by means of the HLCC-based feature. From the experimental results, we have confirmed the feasibility of the proposed interface.

6. Acknowledgements

This work was supported by KAKENHI 20700471, funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of the Japanese Government.

7. References

- [1] G. E. Miller, T. E. Brown, and W. R. Randolph, "Voice controller for wheelchairs," *Medical & Biological Engineering and Computing*, vol. 23, no. 6, pp. 597–600, 1985.
- [2] R. Amori, "VOCOMOTION—an intelligent voice-control system for powered wheelchair," in *Proceedings of the 15th RESNA Annual Conference*, pp. 421–423, Tronto, Canada, 1992.
- [3] W. McGuire, "Voice operated wheelchair using digital signal processing technology," in *Proceedings of the 22nd RESNA Annual Conference*, pp. 364–366, 1999.
- [4] R. C. Simpson and S. P. Levine, "Voice control of a powered wheelchair," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 2, pp. 122–125, 2002.
- [5] A. Sasou and H. Kojima, "Noise robust speech recognition for voice driven wheelchair," in *Proceedings of Interspeech 2007*, pp. 250–253, 2007.
- [6] Y. Matsumoto, T. Ino, T. Ogasawara, "Development of intelligent of wheelchair system with face and gaze based interface," in *Proceedings of IEEE Workshop on Robot and Human Communication*, pp. 262–267, 2001.
- [7] P. Jia, H. Hu, T. Lu, K. Yuan, "Head gesture recognition for hands-free control of in intelligent wheelchair," *J. Industrial Robot*, vol. 34, no. 1, pp. 60–68, 2007.
- [8] B. Raychev, I. Yoda, K. Sakae, "Head pose estimation by nonlinear manifold learning," in *Proceedings of International Conference on Pattern Recognition*, vol. 4, pp. 462–466, 2004.
- [9] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, No. 3, pp. 276–280, 1986.