Acoustic Head Orientation Estimation Applied to Powered Wheelchair Control

Akira Sasou
National Institute of Advanced Industrial Science and Technology, AIST
Central #2, 1-1-1 Umezono Tsukuba Ibaraki Japan, 305-8568
e-mail: a-sasou@aist.go.jp

Abstract— **In this paper, we propose an acoustic-based head orientation estimation method using a microphone array mounted on a wheelchair, and apply it to a novel interface for controlling a powered wheelchair. The proposed interface does not require disabled people to wear any microphones or utter recognizable voice commands. By mounting the microphone array system on the wheelchair, our system can easily distinguish user utterances from other voices without using a speaker identification technique. The proposed interface is also robust to interference from surrounding noise. From the experimental results, we confirm the feasibility and effectiveness of the proposed method.**

Index Terms— **wheelchair, microphone array, noise robust, head orientation, sound source localization**

I. Introduction

Various voice-driven wheelchairs have been developed that try to make it possible for disabled people to move according to their will [1-4]. However, problems with conventional voice-driven wheelchairs still exist. For instance, such wheelchairs use a headset microphone that can record the user's command voice at a higher Signal-to-Noise Ratio (SNR) even with the presence of surrounding noises; however, disabled people need to wear the headset microphone each time they use the voice-driven wheelchair. When the headset microphone moves away from the position of the mouth, disabled people also need to adjust the position of the headset microphone themselves. Because these actions are not always easy for disabled people, we think headset microphones are not practical. In addition, it is not easy for some disabled people to utter voice commands clearly. Such inarticulate commands cause inaccurate speech recognition.

In this paper, we propose a method of acoustic head orientation estimation (AHOE) and a fundamental idea for a novel interface for controlling a powered wheelchair using the AHOE. Our interface does not require disabled people to wear any microphones or utter recognizable voice commands. A microphone array system can estimate the position and/or arrival direction of a sound source even in a noisy environment. When a disabled person utters a sound, the position estimated by the microphone array system mounted on the wheelchair indicates the mouth position, from which we can determine the direction the disabled person is facing. Then our system drives the wheelchair in the estimated direction. Thus, our system does not need to recognize the uttered sound. The proposed interface requires only two capabilities: the ability to face towards a desired direction and to utter an arbitrary sound.

Some researchers developed gaze or head gesture interfaces for powered wheelchair control based on visual information [8-10]. The purpose of this paper is to propose the novel interface for powered wheelchair control which is based on the acoustic head orientation estimation, and to show the feasibility of the fundamental idea.

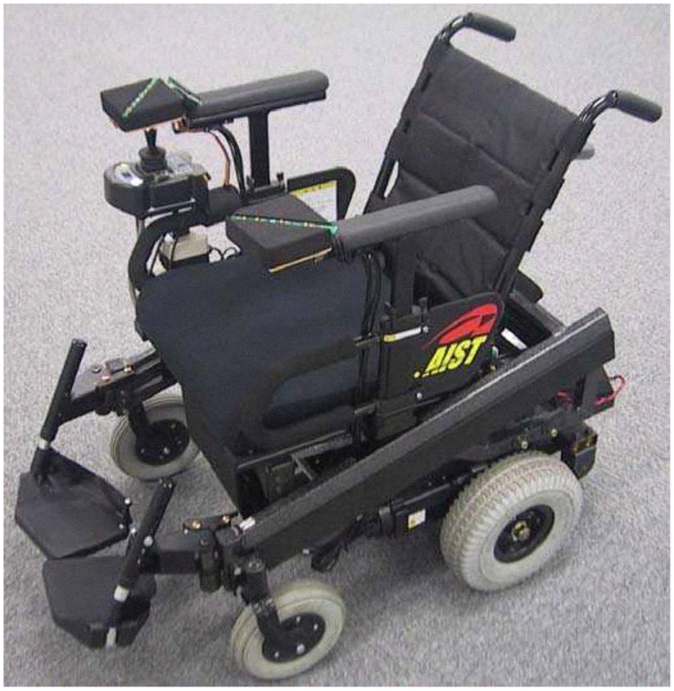

Fig. 1. The developed wheelchair with microphone array system.

II. System Overview

Figure 1 shows the wheelchair we have developed. Our microphone array system consists of two circuit boards. Each circuit board has four silicon microphones soldered every 3 cm linearly. Each circuit board is W130 × D10 × H5 mm in size. The circuit boards are placed along the diagonals of square black sponges. When considering surrounding noise, we would like to put microphones close to the user's mouth as much as

ROBOCOMM 2009, 31st Mar – 2nd Apr 2009, Odense, Denmark.
Copyright © 2011 – 2012 ICST ISBN 978-963-9799-51-6
DOI 10.4108/ICST.ROBOCOMM2009.5815

possible. However, such microphones would be dangerous for some disabled people, for instance, those having cerebral palsy with involuntary movements. So the microphone should be placed far enough from the user's mouth that it does not touch the user's head. Because the black sponges containing the microphone array circuit boards are placed on the edges of the arm rests, the user's head never touches the microphone array system even when there are involuntary movements.

One advantage of using the microphone array is that it can localize sound sources to detect the position of the sound source. Figure 2 shows an example of user utterance localizations. In this example, the user was sitting on the wheelchair and made utterances several times in each head orientation. The blue dots represent the results of user utterance localization when the user was facing forward and tilting slightly forward. The yellow dots represent the results when the user was facing forward and tilting slightly backward. The red dots show the results when the user turned the head to the right, and the light blue dots were those of head orientation to the left. In each head orientation, the localized positions of the utterances are distributed around distinguishable areas. Thus, if we define boundaries between the areas, we can estimate head orientation by identifying the area to which the localized utterance position belongs.

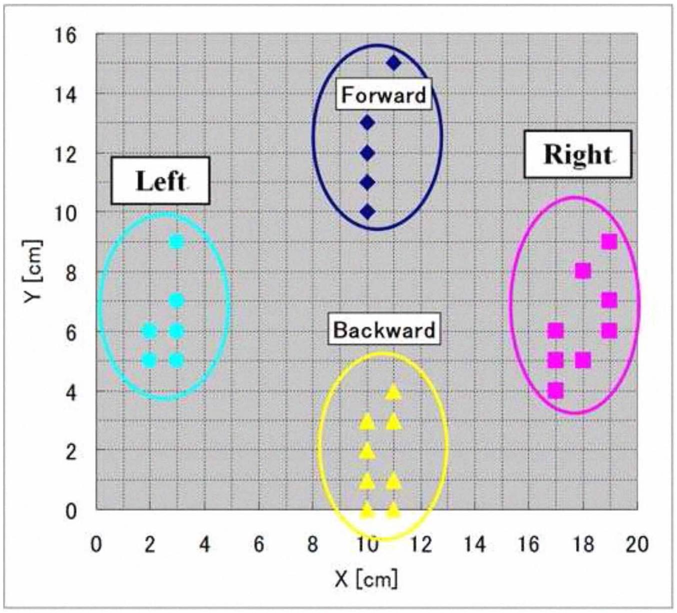

Fig. 2. An example of sound source localization.

The AHOE system should accept only the user's utterances and reject others' utterances and noises from outside the wheelchair. To distinguish the user's utterances from surrounding noises, we have used the positions of sound sources i.e. the mouth position of a seated user is always in a certain area, which is around the centre of the seat at sitting height. We call this the user-utterance-area (UUA). The space of the UUA is set to be large enough to capture all the natural movement of the user's head, and the UUA is restricted so that nobody can be in the UUA except the user. When an observed utterance is judged to be made in the UUA, the AHOE system accepts the utterance as the user's command. On the other hand, when an observed utterance is judged to be made outside the UUA, the AHOE system rejects it. Therefore, by adopting the microphone array system, we can easily distinguish the user's utterances from others' utterances and noises without any training procedures like a speaker identification method.

The processing system consists of one Pentium-M 2.0 GHz CPU board, an 8ch A/D converter and a DC–DC converter. These devices can be put in an aluminium case of W30 × H7 × D18 cm size, which can be hidden under the seat. The system devices that can be seen from the outside are just the microphone array system and LCD showing the status of the system.

III. Sound Source Localization

Human sound radiation depends on frequency and tends to form a directional pattern for higher frequencies with weaker sounds behind the head than in front [5]. In our case, because the microphone array is always in front of the user and never behind, we simply assume that the voice propagates uniformly to the microphone array. Therefore, we currently do not take the radiation pattern into account when estimating head orientation, but use only the sound source position in the UUA.

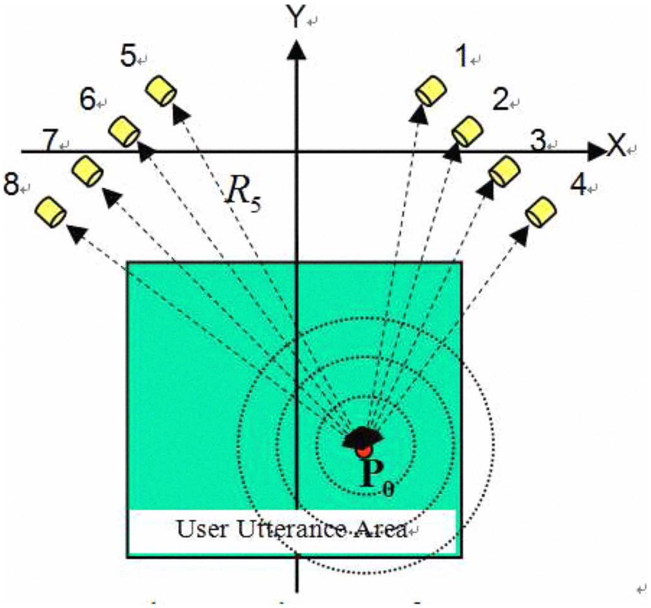

Fig. 3. Schematic diagram of wave propagation.

We have adopted the MUSIC [6] method for estimating the position of sound sources. We assume that a sound source occurring in the UUA is received by the microphone array system as a spherical wave, as shown in Fig.3. The steering vectors in the UUA are defined as follows.

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

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

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

Digital Object Identifier: 10.4108/ICST.ROBOCOMM2009.5815
http://dx.doi.org/10.4108/ICST.ROBOCOMM2009.5815

$$\mathbf{a}(\omega,\mathbf{P}_0)=\left[g_1e^{-j\omega\tau_1},\cdots,g_8e^{-j\omega\tau_8}\right]^T/\sqrt{\sum_{q=1}^{8}g_q^2} \qquad (1)$$

where $\mathbf{P}_0$ is the position of the sound source in the UUA, $\mathbf{P}_1\cdots\mathbf{P}_8$ are the positions of the microphones, τ_q is propagation time, R_q is the distance between the qth microphone and the sound source, v is the sound speed, $\mathbf{a}(\omega,\mathbf{P}_0)$ is the steering vector of user utterance, e is the base of natural logarithms, j is the imaginary number defined as $j^2=-1$, T represents transposition of vector or matrix and $g(\omega,R_q)$ is a distance-gain function. We measured the distance-gain function at several distances and made a model function fitted to the measured values.

The spatial correlation matrix is defined as

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

where $\mathbf{y}(\omega,n)=[Y_1(\omega,n),\cdots,Y_8(\omega,n)]^T$, and $Y_q(\omega,n)$ represent the fast Fourier transform (FFT) of the nth frame received by the qth microphone. The eigenvalue decomposition of $R(\omega)$ is given by

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

where $\mathbf{E}(\omega)$ denotes the eigenvector matrix, which consists of the eigenvectors of $\mathbf{R}(\omega)$ as $\mathbf{E}(\omega)=[\mathbf{e}_1(\omega),\cdots,\mathbf{e}_8(\omega)]$, and $\mathbf{L}(\omega)$ is a diagonal matrix whose diagonal elements consist of the eigen values: $\lambda_1(\omega)\geq\cdots\geq\lambda_8(\omega)$,

$$\mathbf{L}(\omega)=\mathrm{diag}(\lambda_1(\omega),...,\lambda_8(\omega)). \qquad (4)$$

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

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

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.

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

The eigenvectors corresponding to these eigenvalues form the basis of the signal subspace $\mathbf{E}_s(\omega)=[\mathbf{e}_1(\omega),\cdots,\mathbf{e}_{N_{snd}}(\omega)]$. The remaining eigenvectors $\mathbf{E}_n(\omega)=[\mathbf{e}_{N_{snd}+1}(\omega),\cdots,\mathbf{e}_8(\omega)]$ are the basis of the noise subspace.

User utterances are detected according to the following method. First, we search for the position P_0 that absolutely maximizes the following value in the user utterance area.

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

If the absolute maximum value $Q(\mathbf{P}_0)$ exceeds the threshold value T_{usr}, we judge that the user made a sound.

IV. Accuracy Assessment of Sound Source Localization

First, we evaluated the sound source 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 three speakers sitting on the wheelchair and making utterances in a silent room. The speakers were able-bodied, because the purpose of the experiments was to assess the accuracy of sound source localization. Each speaker 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 × 16 cm, as shown in Fig.2, and the sound source localization was executed on a 1 × 1 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. Figure 4 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 in Eqn. (2) due to sound source position changes are smaller along the Y axis than along the X axis.

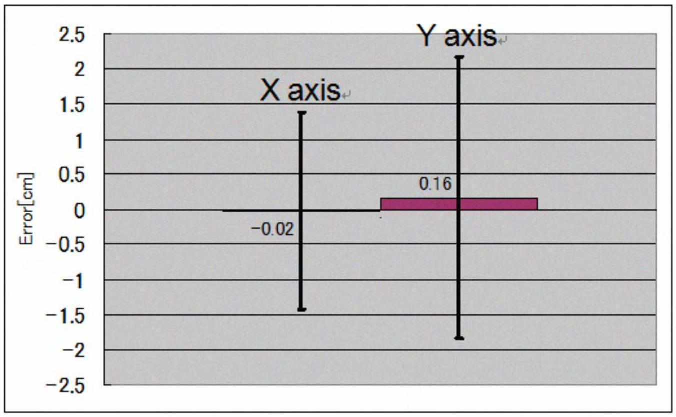

Fig. 4. Evaluation results of sound source localization accuracy.

Next, we evaluated the noise robustness of the sound source 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 5 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 sound source 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. Figures 6 and 7 represent the error distances on the X and Y axes, respectively.

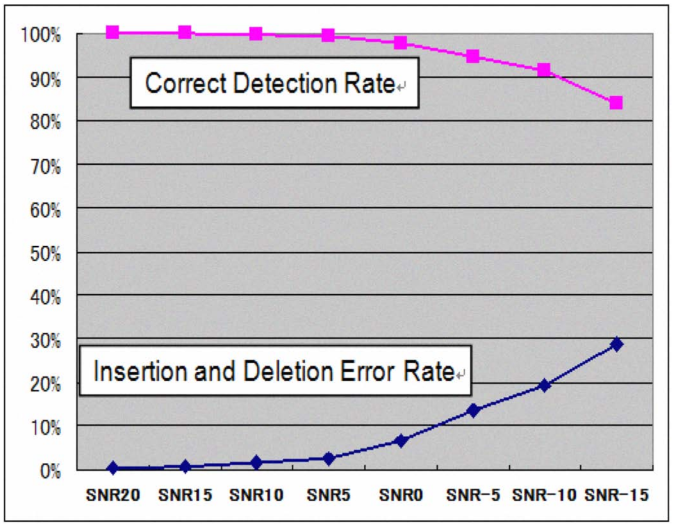

Fig. 5. Evaluation results of utterance detection in noisy environments.

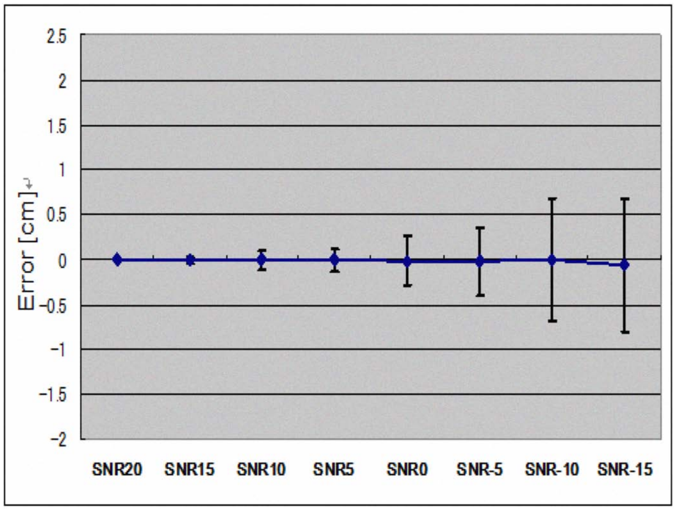

Fig.6. Sound source localization accuracy along the X axis in noisy environments.

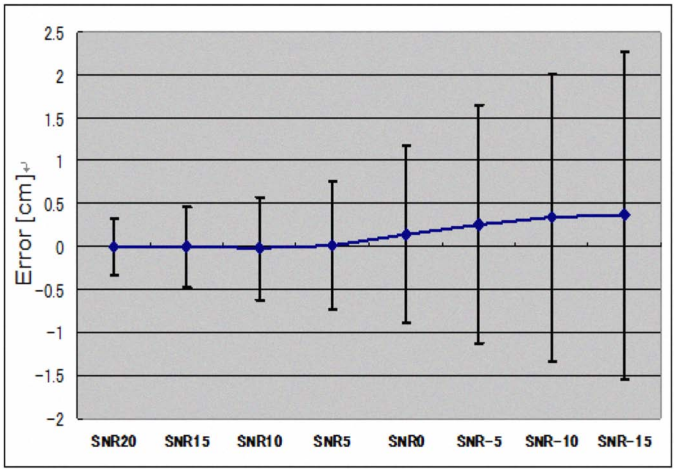

Fig.7. Sound source localization accuracy along the Y axis in noisy environments.

V. WHEELCHAIR CONTROL INTERFACE

We have implemented a preliminary interface for controlling a wheelchair in order to confirm the feasibility of the developed system. Figure 8 represents an assignment of each function to the UUA of size 20 × 16 cm divided into four areas. If we take into account the accuracy of sound source localization examined in the previous section, the space of each area is large enough to be distinguishable from other areas by the developed system.

Figure 9 shows the flow chart of wheelchair control. First, we record N successive frames of speech data, and then calculate the special correlation matrix in Eqn. (2) and execute the sound source localization process. If an utterance can be detected in the UUA, and its duration exceeds a threshold, we then determine the function of wheelchair control from the utterance position in the UUA.

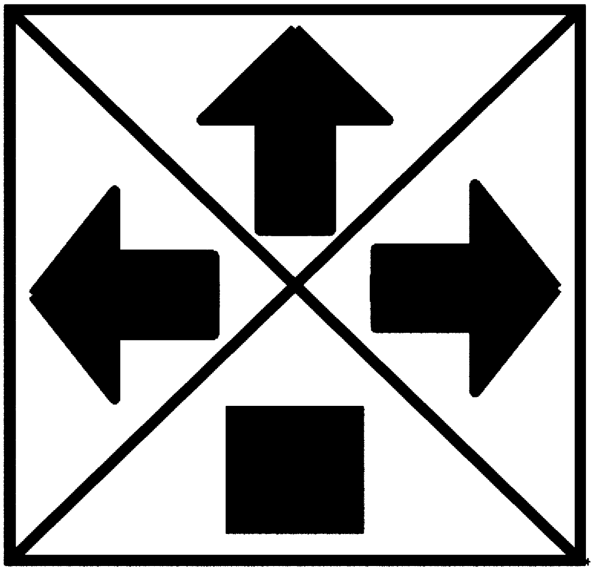

Fig.8. Assignment of each function to the UUA.

Digital Object Identifier: 10.4108/ICST.ROBOCOMM2009.5815
http://dx.doi.org/10.4108/ICST.ROBOCOMM2009.5815

Figures 10–13 represent the wheelchair being controlled by the proposed system in a noisy environment [7]. The loudspeaker behind the wheelchair acts as a noise source, emitting music and a female voice. The user makes sounds by breathing. The wheelchair moves forward when the user makes an utterance while facing forward and tilting slightly forward (Fig.10). The wheelchair moves to the right when the user makes an utterance while facing to the right (Fig.11), and to the left when he does so to the left (Fig.12). The wheelchair stops when the user makes an utterance while tilting slightly backward (Fig.13).

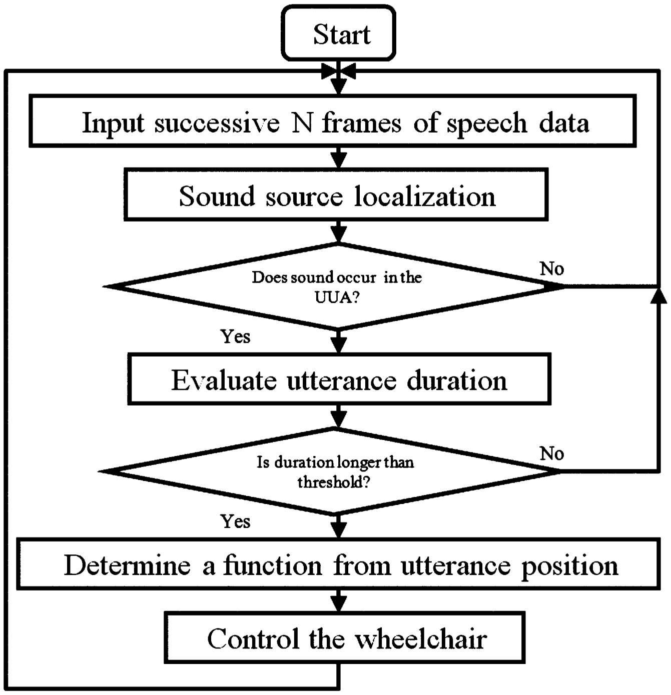
Start
Input successive N frames of speech data
Sound source localization
Does sound occur in the UUA?
No
Yes
Evaluate utterance duration
Is duration longer than threshold?
No
Yes
Determine a function from utterance position
Control the wheelchair

Fig.9. Flow chart of wheelchair control.

VI. Conclusion

In this paper, we proposed an acoustic-based head orientation estimation method using a microphone array mounted on a wheelchair, and tested its application to a novel interface for controlling a powered wheelchair. Our system analyzes the direction in which non-verbal utterances are made. Thus, it does not require the use of a headset microphone or voice identification technique, and it makes fewer demands on the abilities of the user than previous voice-driven wheelchairs. The microphone array system enables identification of the user's utterances while rejecting others' utterances and background noise. We assessed the accuracy of the proposed system's sound source localization and evaluated the system's performance in noisy environments. Our results indicate that the system is feasible and effective for use in controlling the movement of a powered wheelchair in real-life situations.

Acknowledgment

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

References

[1] Miller,G.E., Brown,T.E., Randolph,W.R., "Voice controller for wheelchairs," Med.&Biol.Eng.&Comput., 23, pp597-600, 1985

[2] R. Amori, "VOCOMOTION --- An intelligent voice-control system for powered wheelchair," in Proc. 15th Annu. RESNA Conf. Toronto, Canada, 1992, pp.421-423.

[3] W.McGuire, "Voice operated wheelchair using digital signal processing technology," in Proc. 22nd Annu. RESNA Conf., 1999, pp.364-366.

[4] R.C. Simpson, S.P. Levine, "Voice Control of a Powered Wheelchair," IEEE Trans. Neural Sys. Rehab. Eng., vol.10, No.2, pp.122-125, June 2002.

[5] A. Brutti, M. Omologo, P. Svaizer, "Oriented global coherence field for the estimation of the head orientation in smart room equipped with distributed microphone arrays," in Proc. of Interspeech, pp.2337-2340, 2005.

[6] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag., vol.AP-34, No.3, pp.276-280, March 1986.

[7] The demonstration video is available at the following URL. http://staff.aist.go.jp/a-sasou/wheelchair_ahoe.html

[8] Y.Matsumoto, T.Ino, T.Ogasawara, "Development of intelligent of wheelchair system with face and gaze based interface," Proc of IEEE Workshop on Robot and Human Communication, pp.262-267, 2001.

[9] 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.

[10] B.Raytchev, I.Yoda, K.Sakaue, "Head pose estimation by nonlinear manifold learning," Proc. International Conference on Pattern Recognition, vol.4, pp.462-466, 2004.

Digital Object Identifier: 10.4108/ICST.ROBOCOMM2009.5815
http://dx.doi.org/10.4108/ICST.ROBOCOMM2009.5815

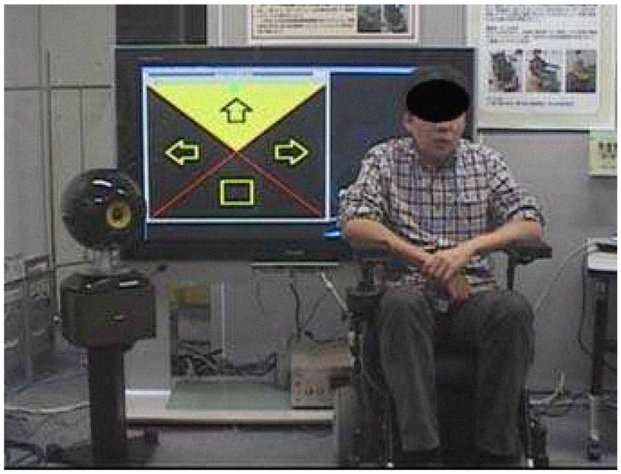

Fig.10. When the user tilts his head forward and makes a sound, the wheelchair goes forward.

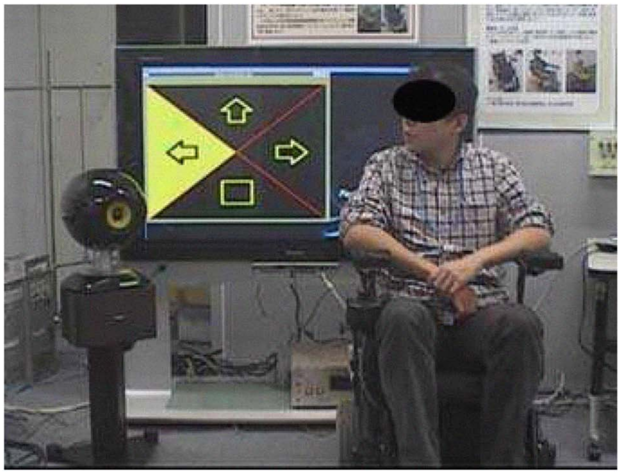

Fig.11. When the user turns his head right and makes a sound, the wheelchair turns right.

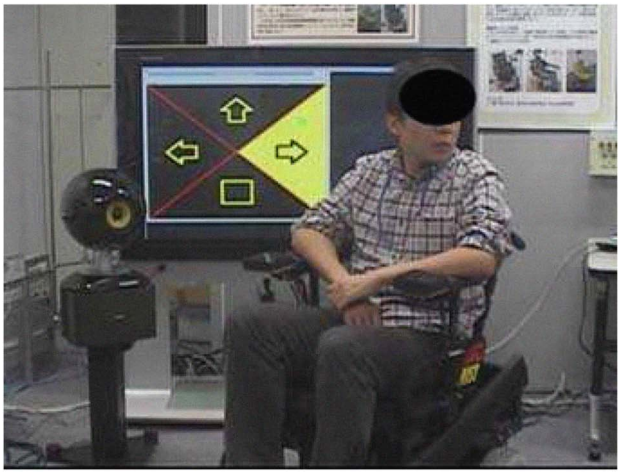

Fig.12. When the user turns his head left and makes a sound, the wheelchair turns left.

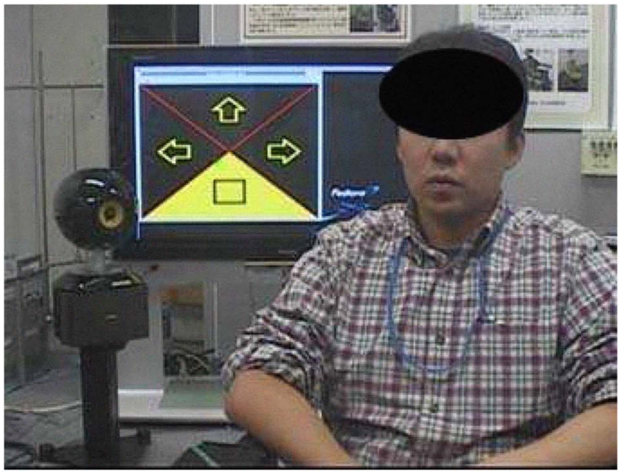

Fig. 13. When the user tilts his head backward and makes a sound, the wheelchair stops.

Digital Object Identifier: 10.4108/ICST.ROBOCOMM2009.5815
http://dx.doi.org/10.4108/ICST.ROBOCOMM2009.5815