

Active Audition System and Humanoid Exterior Design

Kazuhiro Nakadai[†], Tatsuya Matsui[†], Hiroshi G. Okuno^{†*}, and Hiroaki Kitano^{†‡}

[†]Kitano Symbiotic Systems Project, ERATO, Japan Science and Technology Corp.
Mansion 31 Suite 6A, 6-31-15 Jingumae, Shibuya-ku, Tokyo 150-0001, Japan

Tel: +81-3-5468-1661, Fax: +81-3-5468-1664

* Department of Information Sciences, Science University of Tokyo

[‡]Sony Computer Science Laboratories, Inc.

{nakadai, tmatsui}@symbio.jst.go.jp, okuno@nue.org, kitano@csl.sony.co.jp

Abstract

In this paper, we present that a humanoid active audition system improved noise cancellation using humanoid cover acoustics and the cover design for an industrial exterior design.

From the viewpoint of a design, the cover of humanoid should be functional and beautiful. And, the humanoid exterior should not obstruct the motion of the humanoid. In other words, a humanoid with motion requires design for motion. This is an important issue in the robotics field because exterior design of robots has not been sufficiently studied yet.

As for active audition, it is an important problem to distinguish between the internal sounds like motor noises and sounds which originate from the outer world. It is important that inevitable motor noise is cancelled while the humanoid is in motion. Therefore, humanoids require the exterior to cancel such internal noises because it separates humanoid inner world from the outer world.

We reported an active audition system focused on the sound source tracking by integrating audition, vision and motor movements using sound separation of the humanoid cover.

The experiments show that the humanoid can track and localize sound sources more accurately, however noise cannot always be cancelled out optimally and further improvements are required.

Keywords : active audition, humanoid exterior design, noise cancelling, exterior acoustics

1 Introduction

We have been studying humanoids to understand high-level perceptual functions and their multi-modal integration. We use an upper-torso humanoid robot *SIG* as a platform of our research. We believe that the integration of multi-modal sensory input and high degree-of-freedom(DOF) is essential for intelligence [9].

Recently active perception, i.e. the coupling of perception and behavior, has been studied using robots with high DOF [3, 7, 8].

Most of such researches has been carried out in the area of active vision, because it will provide a framework for obtaining necessary additional information by coupling vision with behaviors, such as control of optical parameters or camera positioning using servomotors. For example, an observer controls the geometry parameters of the sensory apparatus to improve the quality of the perceptual processing [1].

On the other hand, in the auditory research field, audition with behaviors, i.e. *active audition*, has not been studied extensively even though people hear sounds while in motion. Indeed, some of robotics researches mention auditory processing with motion, but they assume that the maximum number of sound sources is at most 1 and the input sound is loud enough to ignore motor noises[17, 10]. These assumptions are not enough to understand high-level auditory functions.

People with normal hearing usually can separate sounds from the mixture and focus on a particular voice or sound even in a noisy environment, which is known as the *cocktail party effect*. As traditional auditory research attempts to understand such psychological phenomenon, Computational Auditory Scene Analysis (CASA) studies a general framework of sound

processing and understanding [4, 5, 13, 16]. Its goal is to understand an arbitrary sound mixture including speech, non-speech sounds, and music in various acoustic environment. It requires not only the understanding of the meaning of specific sounds, but it also identifies the spatial relationships of sound sources to further understanding of the environmental sound landscape. This leads to the need of active audition that has capability of dynamically focusing on specific sounds, and active motor control to obtain further information using audition, vision, and other perception.

Active audition creates inevitably complex and irregular motor noises because of humanoid motion. We use microphones which are located near the motors to capture motor sounds which should be ignored in sound processing.

An important idea to process sounds under a noisy environment is the cover of the humanoid because the internal sounds such as motor noises and the sounds originated from the outer world can be distinguished. *SIG* has two pairs of microphones to separate the outer world from the inner world. One pair of microphones are installed at ear position of the head to gather sounds from outer world. The other pair of microphones are installed on the inside of the cover very close to the corresponding microphones to gather sounds from inner world.

In addition, An industrial cover design requires the functionality and beauty. Such a robot industrial design is a new research field.

The paper is organized as follows: In Section 2, humanoid *SIG* and the exterior design are discussed. Section 3 presents the active audition system. Section 4 presents acoustics of the humanoid cover. Section 5 proposes a new sound source localization method by using acoustic measurements. Section 6 shows evaluation of our new localization method, and last two sections give discussion and conclusion.

2 *SIG* And The Exterior Design

We designed *SIG* as a testbed of integration of perceptual information to control motor of high DOF. Its mechanical structure is shown in Fig. 1(a). *SIG* has 4 DOF of body driven by 4 DC motors, a pair of CCD color cameras (Sony EVI-G20) representing the eyes, and two pairs of nondirectional microphones (Sony ECM-77S). The cover, shown in Figure 1(b), is designed as described in the following sections.

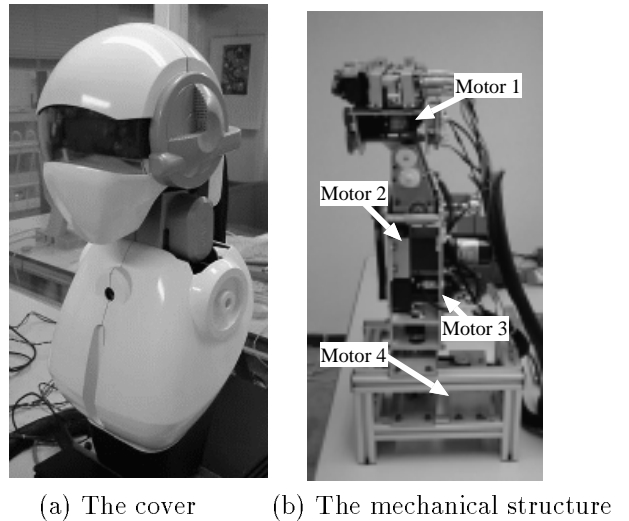


Figure 1: *SIG* the humanoid

2.1 About Robot Design

Robot design. Once the domain of SF movies and books; it will soon enter into the everyday vernacular as recent advancements in artificial intelligence research and robot engineering have edged these technologies out of rarely known research institutes and into the glare of consumer life. The conversion from it's humbler roots in factory automation to that of a coveted consumer product was realized when SONY CORP launched AIBO(ERS-110) in 1999. Early robot prototypes were merely an assemblage of parts and substantial design was not carried out until further advancements in their sensors and mobile parts made its overall design imperative to the accuracy of its functions. Previously, a lack of mobility and autonomy in movement relegated the robot's function most suitable to the enterprise of mass production. A robot with independence of movement and a capability to act autonomously in accordance to its recognition of objects and surroundings has until now been developed for research purposes only. Thus design has played a minimal role at best until these recent advancements in the subtler aspects of robot function has called into question the degree to which design, specifically appearance has on the overall accuracy, particularly in movement. Design is now being recognized as instrument to all aspects of robot development. As robot technology expands into the field of entertainment, the degree to which appearance imparts its desirability as a product is yet another challenge faced by the designer. Similarly, the conversion period of the personal computer (PC) from an object of industry to a

home entertainment unit was accomplished in part by the role design played in promoting its attractiveness as a consumer product. The success of APPLE II best exemplifies the contributions design has made to convert industrial products into general use by a consuming public in 1977. With the humanoid robot under development as our project in mind, we advocate the establishment of a new field called robot design.

2.2 Robot Design As A New Field In The Industrial Design

Industrial design as it exists today, is the design of fundamentally static objects. The structure itself hardly changes. The automobile, an object whose function is movement is still designed as a static object despite having potentially several mobile parts. A door, for example, is intended to be manipulated by the user and not designed with a cognitively autonomous target in mind. The freedom of movement afforded by a robot differs fundamentally from existing industrial static products. The continuance of movement with its plural mobile parts presents several obstacles in pursuit of form amid these restrictions.

As in the case of all industrial products design must be realized amid severely restrictive conditions. The body of a Formula 1 racing car is realized on the basis of aerodynamics, engine installation and so on. These restrictions are even more conspicuous in the case of robot design because a robot has many mobile parts such as a neck, shoulders, elbows, and so on. Hollywood versions of robots like C3PO of the Star Wars saga are visual simulations of imagined mechanical functions, free of these restraints. In the case of the robot, generally the mobile parts serve the same function as the motor and oil pressure system. The highly developed running gear has not yet been developed small enough to accommodate the kind of robot that appears in the work of science fiction. We need to recognize that robot design is not simply the design of an object but the design of a whole range of dynamics.

2.3 What Person Is The Robot?

What is it exactly, or rather, who is it we are designing ultimately comes into question. The first point we have to consider is that the robot is an autonomous instrument. From that point the approach broadens the paradigm of conventional industrial design. Industrial design, we must remember is the design of an instrument. It is the design of objects whether they are household appliances, cars, space shuttles or even most basic tools like the hammer and saw. However,

even these instruments are going through an evolution of their own as we began to produce and use these instruments as tools necessary for our very survival. Once animal bones and stones sufficed for this purpose, however it was the invention in the middle of the 20th century of the digital computer that was fundamentally different from conventional instruments. The computer was the first instrument to operate information in the place of the object; whose function was not physics but rather the abstractor concept of programming. Although this has afforded a certain DOF for the designer it has also meant simultaneous chaos, resulting from the separation of the one that performs the function from the idea which has occurred. Design must be approached from the view point of how it can serve as a hand in enabling clear, easy access to the user and put into context, the assemblage of chaos within. In his book "the Psychology of Everyday Things" D.A.Norman examines the role of interface design in communication the function of the object. (and it's desirability as a product.) As much as design had to assume and be equal to the complications that arose with the appearance of the computer, so it must again broaden its role in the genesis of robot engineering and technology [15]. The autonomous nature of the robot creates the division between itself and more conventional instruments. Interaction between human and present robots is still far from satisfactory, however the acceleration of research in this field makes it a matter of time before the interactive possibilities of the robot will separate it finely from mere instruments. What is inside the robot will be largely influenced by its appearance. What kind of relation it will have to its structure is a theme that has to be explored. This is a problem particularly applicable to the non-humanoid robot. An example of this kind can be shown in the development of AIBO. Those who participated in the SONY CORP project from beginning, first visualized a kind of insect, as six legs facilitate the smoothness of movement. But since the first prototypes conveyed less an insect than the universally reviled cockroach, a substantial design change was necessary. While the four-legged creature lacked smoothness in movement, it was not however, a liability. In fact, the awkward, rambling gait of the earliest prototype was decisive in SONY's decision to support the project. Robot development in this field, it was decided cannot be viewed as the development of alternate animals but rather as robots having similar attributes but functions entirely separate from those of living pets. Although AIBO is often called a pet robot it is more accurately in the realm of digital creatures

as reflected in its design which does not shy away from its mechanical functions. On the contrary, digital creatures are by design a sum of their mechanical parts. The same is true of the time the humanoid robot was introduced into the Kitano project. During research into human intelligence the first humanoid would have had the appearance of a human through the applications of silicone and cosmetically enhanced features. The awareness however, that project members were pioneering a whole new field of industrial design was instrumental in rejecting non-essential visual aids on the prototype. This would have gone against the spirit of the project, which was again not replicating a human being but analyzing the structure of a human's cognitive functions.

2.4 Humanoid Robot

It is the main goal of this project to explore the basic theory of human intelligence. Therefore the visual and auditory capabilities of the autonomous robot must be equipped with visual and auditory features aligned much the same way human's are in order to assess their integration into our cognitive functions. It was discovered that the processing of external voices became very difficult, making it necessary to attach a cover over the auditory mechanism of the robot to block the sound from outside. This combined with the sound of the motor which occurs inside the robot-another interference, was decisive in introducing an element of design. It is to understand the intelligence of the human being that makes the humanoid a necessary element in our research, however, the nature of the project cannot be expressed sufficiently by the mere replication of human. The structure of the robot being similarly adapted to the human form is one option towards efficiency. Design plays an important role in opening up a whole new field of inquiry; one which all researchers can value as a visually expressive validation of their work.

2.5 Experiment Environment And Proportion Setting

This particular robot will remain forever consigned to the laboratory for experimental purposes. Here the robot's dimensions will signify the entertainment humanoid robot's place in the consumer's home. At present, the research robot sits in 12 square meter room crowded in by two desks and at any given time two or three researchers. The entertainment robot will therefore have to adapt in much the same way to its surroundings. Its size will be an important considera-

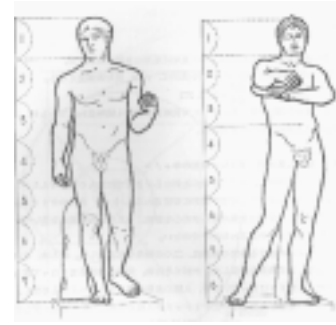


Figure 2: Difference of the impression about eight heads high and seven heads high

tion when consumers decide whether or not to allow it into their homes and will most likely reject it if its dimensions threaten to overpower the space it is meant to share. The robot must settle as compactly as possible. This is particularly relevant to Japanese houses, with an average living room size of 12 square meters. The impression of distance is often measured subconsciously. A robot's proportions however, must adapt to its height. Therefore, a compact, child sized robot uncomfortably conveys its function as a toy so whereupon a small system is desirable it must not obfuscate its meaning by turning it into something entirely unintended. We have judged that an adult woman, has the dimensions desirable for this purpose and can accommodate the small systems designs that already exist for this purpose. Although small in stature, it suggests the form of a human being ideal for use in limited space. The research robot used in this project has been designed with arm sockets despite not having any mobility in the body to entice the imagination into forming a complete picture; a playful homage to the Venus de Milo whose very lack is at the center of her embodiment of complete beauty. This is how design is succeeding as an interactive medium between the viewer and the object. The imagined product at eight heads high has its roots in classical Greek sculptures. Ideals of feminine grace, dating back as early as the 5th century B.C. have its roots in the Golden module which determines the 3:5:8 configuration as the fundamental basis by which the body is measured and proportioned (see also Figure 2).

If the face is measured vertically at 21 cm and again from the chin to the top of the breast down to the navel, these dimensions multiplied by eight comprise the ideal by which beauty is measured and by which we can determine the robot's dimensions as an ideal.

This is how we formulated a concept by which our aesthetic standard evolved. The exterior design of this humanoid robot will further facilitate research to a higher degree than before by putting into visual and human context the full range of dynamics at work. Fig 3, 4, 5 and 6 show processes of *SIG* exterior.



Figure 3: Process of product 1 - sketching -

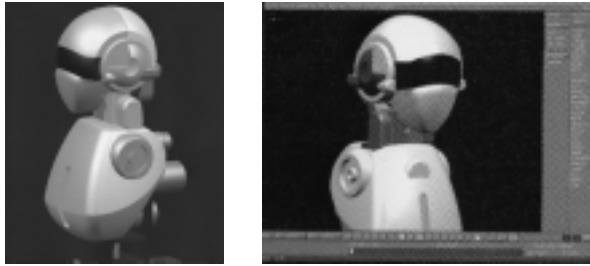


Figure 4: Process of product 2 - 3D modeling for mold production -



Figure 5: Process of product 3 - model production -



Figure 6: Process of product 4 - final coating -

3 Active Audition System

The active audition system captures sounds at 48 KHz sampling rate with two pairs of microphones and

transforms the sounds to sound spectrograms by FFT for 1,024 points. Then, the sound spectrograms are processed by two components; internal sound suppression, and sound stream separation.

3.1 Internal Sound Suppression

Motor noises are cancelled by applying a kind of adaptive filter on the input sound spectrograms created by FFT. The filter uses *heuristics with internal microphones*, which specifies the condition to cut off burst noise mainly caused by motors. The heuristics orders that localization by sound or direction-pass filter ignore a subband if the following conditions hold:

1. The power of internal sounds is much stronger than that of external sounds.
2. Twenty adjacent subbands have strong power.
3. A motor command is being processed.

3.2 Sound Stream Separation

We use a new direction-pass filter with direction which is calculated by epipolar geometry.

3.2.1 Localization By Epipolar Geometry

Direction information of sound sources is extracted using auditory epipolar geometry[11]. This idea expands epipolar geometry in vision[6] to auditory field as shown in Fig 7. This method extracted direction information without using *Head Related Transfer Function(HRTF)*. It is useful to localize sound source without using *HRTF* because it is easy to change if surrounded environment is changed. So, *HRTF* is hard to use for sound source localization in real environments.

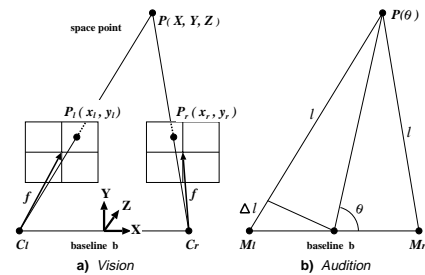


Figure 7: Epipolar geometry for localization (C_l, C_r : camera center, M_l, M_r : microphone center)

It extracts peaks by using FFT for each subband, 47Hz in our implementation, and calculates the *interaural phase difference(IPD)*. The sound source direction is estimated by Equation (1):

$$\cos \theta = \frac{v}{2\pi f b} \Delta \varphi \quad (1)$$

where v is the velocity of sound, b is the distance (baseline) between left and right microphones, $\Delta\varphi$ is IPD and f is the frequency of sound. For the moment, we assume the velocity of sound is fixed to 340m/sec and is invariant to temperature and humidity.

By using the relative position between camera centers and microphones, it is easy to convert from epipolar plane of vision to that of audition (see Figure 7b). In *SIG*, the baselines for vision and audition are in parallel. Therefore, whenever sound source is localized by epipolar geometry in vision, it can be converted easily into the angle θ . This can apply to a method of integration visual and auditory information, and we reported such an integration system of perceptual information based on epipolar-geometry [11].

3.3 Pitch Extraction

Pitches are extracted by a kind of spectral subtraction [2]. It uses peak approximation method based on characteristics of FFT and window function.

Consider that the peak $[\omega_2, y_2]$ is detected, and the values of both neighbors are $[\omega_1, y_1]$ and $[\omega_3, y_3]$ as shown in Fig.8. Then, the true peak $[\omega_0, y_0]$ is estimated as follows:

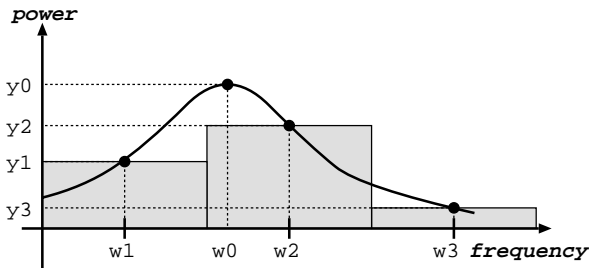


Figure 8: An example of a spectral peak by Fourier Transformation

First, ω_0 is estimated as the following Equation (2).

$$\omega_0 = \begin{cases} \omega_2 + \frac{2\pi (2|y_1| - |y_2|)}{T (|y_1| + |y_2|)} & (\omega_1 < \omega_0 \leq \omega_2) \\ \omega_2 - \frac{2\pi (-|y_2| + 2|y_3|)}{T (|y_2| + |y_3|)} & (\omega_2 < \omega_0 < \omega_3) \end{cases} \quad (2)$$

$$\begin{aligned} Arg(y_0) &= \tan^{-1} \left(\frac{\Im[y_0]}{\Re[y_0]} \right) \\ &= \tan^{-1} \left(\frac{\Im[y_2]}{\Re[y_2]} \right) + \frac{T}{2} (\omega_2 - \omega_0) \end{aligned} \quad (3)$$

$$\begin{aligned} |y_0| &= \frac{\Delta\omega (-T^2 \Delta\omega^2 + 4\pi^2)}{2\pi^2 \sin \frac{T}{2} \Delta\omega} |y_2|, \\ \Delta\omega &= \omega_2 - \omega_0 \end{aligned} \quad (4)$$

And the phase and amplitude of the true peak y_0 are estimated as Equations (3) and (4), respectively.

Because the above equations require a relatively small number of calculations, our method can run faster and extract more accurate pitches. For example, in comparison with Bi-HBSS [14], which is known as a sound source separation system using pitch extraction by spectral subtraction, the proposed method needs only 1/200 of amount of calculation per a peak [12].

3.4 Direction-pass Filter

The direction-pass filter selects subbands that satisfies the IPD of the specified direction. The detailed algorithm is describes as follows:

1. The specified direction θ is converted to $\Delta\varphi$ for each subband (47 Hz).
2. Extract peaks and calculated IPD , $\Delta\varphi'$.
3. If IPD satisfies the specified condition, namely, $\Delta\varphi' = \Delta\varphi$, then collect the subband.
4. Construct a wave consisting of collected subbands.

3.5 Problem In Active Audition System

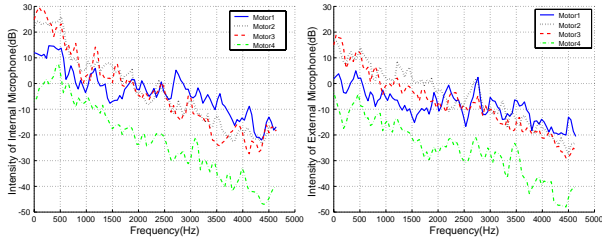
The system, however, has a problem that noise cancellation can not be sufficient because the internal microphones can capture louder sounds originating from the outer world than the external microphones.

We considered that the problem was caused by resonance inside the cover. We needed to measure acoustics of the cover to confirm it and to improve noise cancellation. Acoustic measurement is described as the next section.

4 Acoustic Analysis Of The Cover

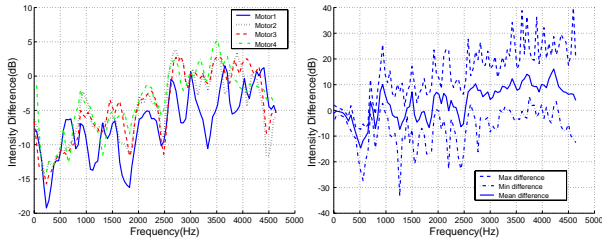
The cover acoustics is measured in an anechoic room. The items of acoustic measurements are shown in the following.

- Frequency response of each motor noise by both internal and external microphones (Figures 9(a) and (b))
- Intensity difference between internal and external microphones. Figure 10(a) shows intensity difference of each motor noise. The graph is estimated by subtracting internal microphone's frequency response from external one. Figure 10(b) shows intensity difference of the outer sounds. This is estimated by impulse responses. The impulse responses are measured at 12 points which are elements of a matrix of horizontal and vertical directions; horizontal directions are $0^\circ, \pm 45^\circ$,



(a) internal microphone (b) external microphone

Figure 9: Frequency Response of Motor Noise



(a) motor noise (b) outer sounds

Figure 10: Intensity Difference

$\pm 90^\circ$ and 180° from robot center and vertical directions are 0° and 30° .

From the figures, main observations are summarized as follows:

1. Motor noise is broadband and is captured less than 30 dB by internal microphones, is captured less than 20 dB by external ones as shown in Figures 9(a) and (b).
2. Motor noise is captured louder by external microphones than by internal microphones where frequency is more than 2.5 KHz as shown in Figure 10(a). This shows that the cover makes easier to capture motor noise by internal microphones, because sounds from outer world is cut off by the cover.
3. Acoustic signals are often captured louder by internal microphones than by external microphones for frequency of less than 2 KHz. Especially, the tendency is more remarkable for frequency of less than 700Hz as shown in Figure 10(b). This shows resonance in the cover. The diameter of the cover is about 18 cm, which is corresponded to $\lambda/4$ at frequency of 500Hz. This causes resonance which has 500 Hz of center frequency. The similar resonance is occurred in Figure 10(a)
4. Internal sound is captured about 10 dB louder than external sound on average by comparing Figure 10(a) and (b). Therefore, the cover effectiveness to separate the inner and outer sounds is about 10 dB.

5 New Noise Cancellation Method

The acoustic experiments shows resonance is occurred near frequency of 500Hz. The resonance is caused by head form because the wavelength of 500Hz corresponds to $\frac{1}{4}$ of the head diameter.

Generally, wide space which is more than a wave length is needed to absorb the wave. For this reason, the resonance of 500Hz, which is relatively long wave length, can not be cancelled even if soft materials for absorption are stuck inside the cover. Then, we revised the noise cancel method by using acoustic experiments as noise models.

Our new method to improve noise cancellation is shown as follows:

1. Intensity difference between external and internal microphones is similar to measured motor noise intensity differences.
2. Intensity and pattern of the spectrum are similar to measured motor noise frequency responses
3. A motor command is being processed.

6 Experiments

In this section, we will demonstrate effectiveness of noise cancellation by new method in sound source localization.

There are two sound sources: two B&W Nutilus 805 loud speakers located in a room of 10 square meters. The room the system is installed in a conventional residential apartment facing a road with busy traffic, and exposed to various daily life noise. Sound environment is not at controlled for experiment to ensure feasibility of the approach in daily life.

One sound source *A* (Speaker A) plays a monotone sound of 500 Hz. The other sound source *B* (Speaker

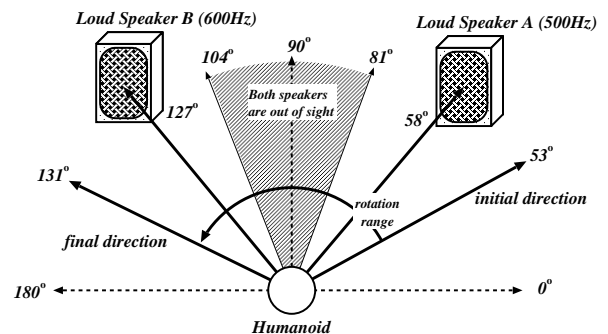


Figure 11: Experiment: Sound source localization while *SIG* moves.

B) plays a monotone sound of 600 Hz. *A* and *B* are located in front of *SIG*. *SIG* turns toward the direction of the sound source *B* using the direction obtained by audition under the condition that both *A* and *B* make sounds. Figure 11 shows this situation.

Figure 12(a) shows the captured spectrogram of *SIG*, Figures 12(b), (c) and (d) show the localization results. The Y axis of each graph describes direction of *A* and *B* in humanoid coordinate system. Figures 12(b), (c) and (d) show the results of sound source localization without noise cancellation, with noise cancellation by the previous method, and with our new noise cancellation, respectively.

In Figure 12(a), we can see 4 burst noises at 5.5, 7.0, 8.1 and 9.0 seconds. And, Figure 12(b) shows that sound source localization is badly impaired. Using our previous method, burst noise at 5.5 and 7.0 seconds are cancelled or weakened as shown in Figure 12(c), but other noises still remain. Figure 12(d) shows that our new method cancels all burst noises and suppresses vibration.

However, location error of $\pm 10^\circ$ can be observed, i.e. when the robot is rotated by 80° , the actual sound source is located at an angle of 100° .

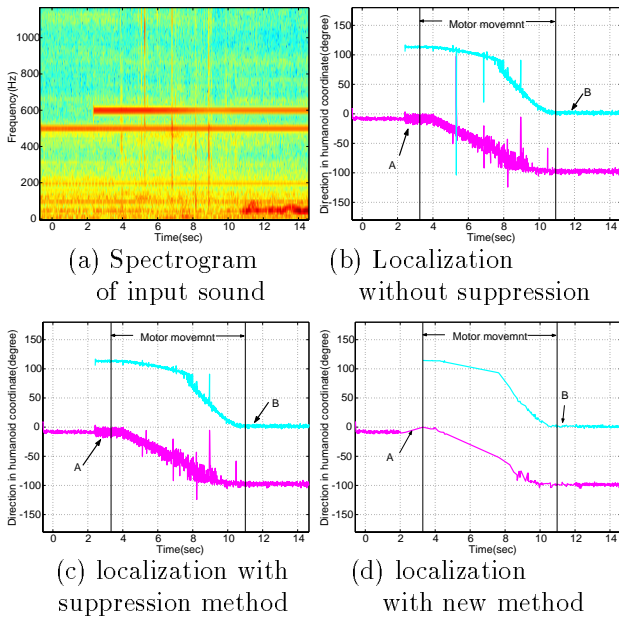


Figure 12: Localization Experiments of sound sources

It is considered that the rotation error is caused by *IPD* errors which is generated on *IPD* calculation. Because the FFT window length and its discreteness contribute to the *IPD* errors, it is difficult to cancel the errors. Other sensory information such as vision and potentiometer outputs can be used to compensate the error.

7 Discussion

At the first design of noise cancellation method, we simply considered that the cover reduced motor sounds emitted to the outer environments. However, the first design did not work well due to resonance within the cover, which we did not realize at the time of design. We solved this resonance problem by a new cancellation method based on the templates of motor noises since it is impossible to eliminate such a resonance completely. One of future work for improving the performance of active audition is to design a robot structure and its cover so that the resonance is reduced as much as possible, in particular, at the frequency range between 20Hz and 20KHz. This range of frequency is important for human communication by using voice. The requirements for better robot design are as follows:

- robot exterior design
 - The inner space is so narrow that resonance occurs only at higher frequencies which humans cannot hear
 - The cover should isolate the inner space completely from the outer space. There is no crack between cover parts.
 - The cover material should block and absorb sounds much more effectively.
- robot mechanical design
 - Much more silent motor drives such as harmonic drives should be used.
 - The number of sound making components such as stoppers and springs should be reduced.

SIG has currently the beautiful and functional cover from the viewpoint of industrial design. Moreover, the cover has 10dB sound blocking effect and relatively simple face form to be able to reduce acoustic calculation. But it is not isolated completely because of some cracks and has resonance at relatively low frequencies. The acoustics improvement of *SIG* cover for active audition with keeping beauty and functionality for industrial design is one of the most important future works.

8 Conclusion

We discuss that it is not only important for industrial design but also for active audition. By using acoustic measurement of the cover, we confirm that our cover design contributes to improve signal-noise ratio of 10dB by separating the inner and outer

world. And we demonstrate an effective noise cancelling method which improves sound source localization even if the humanoid is moving. We also propose acoustic properties of a robot and the cover for better cover design.

We discuss the importance of this research with respect to active audition even though it has not been studied so far. We show localization by epipolar geometry without using *HRTF*. In addition, this method can easily be expanded to combine visual information. It is an important method for the integration of perceptual information as well as to understand fundamental principles of intelligence.

Acknowledgments

We thank *NITTOBO Acoustic Engineering Co., Ltd.* for the acoustic measurements and offering the anechoic room. We also thank our colleagues of Symbiotic Intelligence Group, Kitano Symbiotic Systems Project; Dr. Theo Sabish, Dr. Tino Lourence, Yukiko Nakagawa and Dr. Iris Fermin for their discussion.

References

- [1] Y. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1(4):333–356, 1987.
- [2] S. F. Boll. A spectral subtraction algorithm for suppression of acoustic noise in speech. In *Proceedings of 1979 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-79)*, pages 200–203. IEEE, 1979.
- [3] R. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, and M. Williamson. The cog project: Building a humanoid robot. Technical report, MIT, 1999.
- [4] G. J. Brown. *Computational auditory scene analysis: A representational approach*. PhD thesis, Dept. of Computer Science, University of Sheffield, 1992.
- [5] M. P. Cooke, G. J. Brown, M. Crawford, and P. Green. Computational auditory scene analysis: Listening to several things at once. *Endeavour*, 17(4):186–190, 1993.
- [6] O. D. Faugeras. *Three Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, MA., 1993.
- [7] M. Kawato. Bi-directional theory approach to consciousness. In *Cognition, Computation, and Consciousness*. Oxford University Press, 1996.
- [8] N. Kita, S. Rougeaux, Y. Kuniyoshi, and S. Sakane. Real-time binocular tracking based on virtual horopter. *Journal of Robotics Society Japan*, 13(5):101–108, 1995.
- [9] H. Kitano, H. G. Okuno, K. Nakadai, I. Fermin, T. Sabish, Y. Nakagawa, and T. Matsui. Designing a humanoid head for robocup challenge. In *Proceedings of 4th International Conference on Autonomous Agents (Agents 2000)*. ACM, 2000.
- [10] Y. Matsusaka, T. Tojo, S. Kuota, K. Furukawa, D. Tamiya, K. Hayata, Y. Nakano, and T. Kobayashi. Multi-person conversation via multi-modal interface – a robot who communicates with multi-user. In *Proceedings of Eurospeech*, pages 1723–1726. ESCA, 1999.
- [11] K. Nakadai, T. Lourens, H. G. Okuno, and H. Kitano. Active audition for humanoid. In *Proceedings of 17th National Conference on Artificial Intelligence (AAAI-2000)*. AAAI, 2000. (to appear).
- [12] K. Nakadai, H. G. Okuno, and H. Kitano. A method of peak extraction and its evaluation for humanoid. In *SIG-Challenge-99-7*, pages 53–60. JSAI, 1999.
- [13] T. Nakatani, H. G. Okuno, and T. Kawabata. Auditory stream segregation in auditory scene analysis with a multi-agent system. In *Proceedings of 12th National Conference on Artificial Intelligence (AAAI-94)*, pages 100–107. AAAI, 1994.
- [14] T. Nakatani, H. G. Okuno, and T. Kawabata. Residue-driven architecture for computational auditory scene analysis. In *Proceedings of 14th International Joint Conference on Artificial Intelligence (IJCAI-95)*, volume 1, pages 165–172. AAAI, 1995.
- [15] D. A. Norman. *The psychology of everyday thing*. Basic Books Inc., N.Y., 1988.
- [16] D. Rosenthal and H. G. Okuno, editors. *Computational Auditory Scene Analysis*. Lawrence Erlbaum Associates, NJ., 1998.
- [17] A. Takanishi, S. Masukawa, Y. Mori, and T. Ogawa. Development of an anthropomorphic auditory robot that localizes a sound direction (*in japanese*). *Bulletin of the Centre for Informatics*, 20:24–32, 1995.