

More concretely, first, we use the L1-norm method and obtain tentatively separated sounds. Second, we extract the harmonic structure from these sounds. Finally, we separate the speech mixtures again, with the constraint that the harmonic structure is always powerful. The experiment revealed that our proposed method improved the correctness of ASR by about four points compared to the baseline method.

In future work, we intend to add new constraints for voiceless consonants, which do not have a harmonic structure. Since voiceless consonants have high power in the high-frequency region, we can expect that new constraints will improve the high-frequency region, which our proposed method cannot do. Another area we need to tackle is the reverberation environment. Even though our experiments were carried out using impulse responses in an anechoic chamber, developing a method that works properly in a standard reverberation room is essential for robots that work in real environments.

In addition, we need to take into consideration the speech-recognition module. We expect that a recognition system that can place more stress on the harmonic structure can improve the ASR results since our proposed method could separate the harmonic structure very well. Another way is to use missing feature theory [15], which would enable us to reduce the effect of interference or enable us to recover feature values that are distorted.

ACKNOWLEDGMENTS: Part of this study was supported by a Grant-in-Aid for Scientific Research (S) and the Global COE Program.

REFERENCES

- [1] M. Hirose and K. Ogawa. Honda humanoid robots development. *Philosophical Transactions A*, Vol. 365, No. 1850, p. 11, 2007.
- [2] K. Akachi, G. Miyamori, T. Kawasaki, M. Ishizaki, T. Kimura, T. Isozumi, K. Kaneko, F. Kanehiro, H. Hirukawa, and H. Hasunuma. The development of a humanoid robot: HRP-3. *Nippon Robotto Gakkai Gakujutsu Koenkai Yokoshu (CD-ROM)*, Vol. 24, p. 3H15, 2006.
- [3] K.F. MacDorman and H. Ishiguro. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, Vol. 7, No. 3, pp. 297–337, 2006.
- [4] A. Hyvärinen and E. Oja. Independent Component Analysis: Algorithms and Applications. *Neural Networks*, Vol. 13, No. 4-5, pp. 411–430, 2000.
- [5] L. Griffiths and C. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE Transactions on Antennas and Propagation*, Vol. 30, No. 1, pp. 27–34, 1982.
- [6] O. Yılmaz and S. Rickard. Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions on Signal Processing*, Vol. 52, No. 7, pp. 1830–1847, 2004.
- [7] K. Nakadai, H.G. Okuno, and H. Kitano. Auditory fovea based speech separation and its application to dialog system. In *IEEE/RSJ International Conference on Intelligent Robots and System, 2002*, Vol. 2, 2002.
- [8] T.W. Lee, M.S. Lewicki, M. Girolami, T.J. Sejnowski, et al. Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Signal Processing Letters*, Vol. 6, No. 4, pp. 87–90, 1999.
- [9] P. Bofill and M. Zibulevsky. Underdetermined blind source separation using sparse representations. *Signal processing*, Vol. 81, No. 11, pp. 2353–2362, 2001.
- [10] W. Zhang, J. Liu, J. Sun, and S. Bai. A New Two-Stage Approach to Underdetermined Blind Source Separation using Sparse Representation. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, Vol. 3, 2007.
- [11] Y. Li, A. Cichocki, and S. Amari. Analysis of sparse representation and blind source separation. *Neural Computation*, Vol. 16, No. 6, pp. 1193–1234, 2004.
- [12] Y. Li, S. Amari, A. Cichocki, D.W.C. Ho, and S. Xie. Underdetermined blind source separation based on sparse representation. *IEEE Transactions on Signal Processing*, Vol. 54, No. 2, pp. 423–437, 2006.
- [13] S. Davis and P. Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 28, No. 4, pp. 357–366, 1980.
- [14] S. Winter, H. Sawada, and S. Makino. On real and complex valued L1-norm minimization for overcomplete blind source separation. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 86–89, 2005.
- [15] B. Raj and R.M. Stern. Missing-feature approaches in speech recognition. *IEEE Signal Processing Magazine*, Vol. 22, No. 5, pp. 101–116, 2005.