Intelligent Ubiquitous Sensor Network for Sound Acquisition

Koji Kugata, Tomoya Takagi, Hiroki Noguchi, Masahiko Yoshimoto, and Hiroshi Kawaguchi

Graduate School of Engineering, Kobe University
1-1 Rokkodai, Nada, Kobe, Hyogo, 657-8501 Japan
kugata@cs28.cs.kobe-u.ac.jp

***Abstract*—We propose a microphone array network that realizes ubiquitous sound acquisition. Nodes with 16 microphones are connected to form a large sound acquisition system that carries out voice activity detection (VAD), sound source localization and sound source separation. The three operations are distributed among nodes using network. Because the VAD is implemented to manage power consumption, the system consumes little power when speech is not active. The power of the VAD module is only 2.1 mW on an FPGA. The system can improve an SNR by 7.75 dB using 112 microphones.**

***Index Terms*—Microphone array, ubiquitous sensing, sensor network, distribution network, low-power system**

***ISCAS Track Selection*—Digital Signal Processing**

1. INTRODUCTION

In recent years, information processing technology improvements have realized real-time sound processing systems with microphone arrays. A microphone array can localize sound sources and separate multiple sources using the acquired sounds' spatial information. The computational effort of these operations increases polynomially with the number of microphones, but the performance of these operations is known to improve concomitantly [1]. To reduce the increasing power of a microphone array and to satisfy the recent demand for ubiquitous sound acquisition, it is necessary to realize a low-power, large sound-processing system.

Huge microphone arrays have been widely researched at Tokyo University of Science (128 ch) [2], the University of Electro-Communication (156 ch) [3], Brown University and Rutgers University (512 ch) [4]–[5], and the Massachusetts Institute of Technology (1,020 ch) [1]. However, obstacles to their practical use persist: increasing computation and power consumption.

To implement a microphone array as an actual ubiquitous sound acquisition system, we propose to divide the huge array into sub-arrays and produce a network: an intelligent ubiquitous sensor network (IUSN). The sub-array nodes can be set up on a room's walls and ceiling. The performance can be improved by increasing the nodes, but the communication between nodes does not increase so much in our system.

This research has been supported by the Semiconductor Technology Academic Research Center (STARC).

2. INTELLIGENT UBIQUITOUS SENSOR NETWORK

Fig. 1 presents a brief description of the proposed IUSN and a functional block diagram of a sub-array node. In all, 16 microphone inputs are digitized with A/D converters; the sound information is stored in SRAM. They are then used for sound source localization (SSL) and sound source separation (SSS). The power manager and voice activity detection (VAD) module deactivate the sound processing unit to conserve power: the sound processing unit can be turned off if no sound exists around the microphone array. That power management is necessary because numerous microphones waste much power when not used.

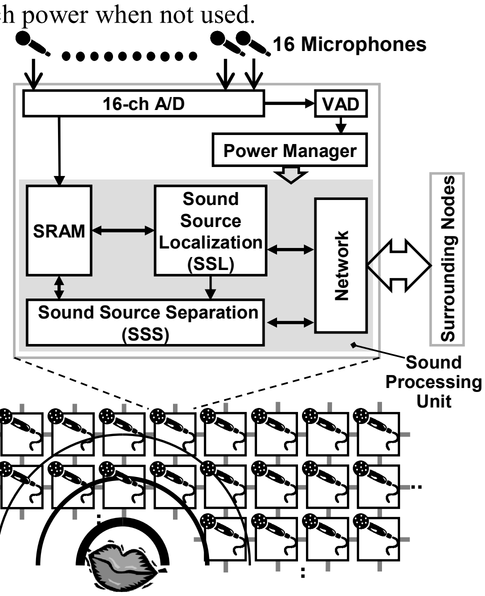

Fig. 1. Intelligent ubiquitous sensor network (IUSN) and block diagram of a sub-array node.

Fig. 2 depicts a flow chart of our system. The salient features of the system are: 1) low-power VAD to activate the entire node, 2) SSL to find sound sources, and 3) SSS to enhance the sound. The sub-array nodes are mutually connected to support their communication. Therefore, the sound gained by each node can be gathered to improve the sound source's SNR further. The system can be characterized as a large microphone array by cooperating with surrounding nodes. Computations can be distributed among nodes. Each node preprocesses acquired sound data. Then only compressed data—localization information and separated sound—are communicated. The system provides scalability in terms of the number of microphones.

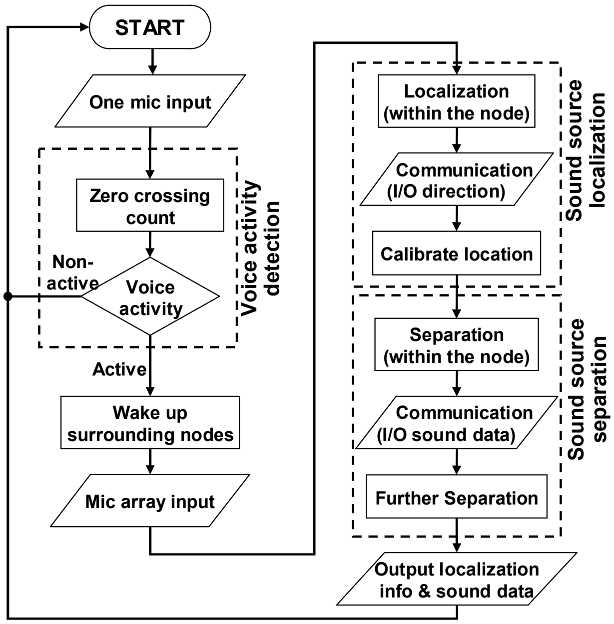

Fig. 2. Flow chart of an intelligent ubiquitous sensor node.

3. ALGORITHMS

· *Voice Activity Detection (VAD)*

A microphone array comprises numerous microphones, which would easily consume much power. Therefore, our intelligent ubiquitous sensor node must operate with a limited energy source and conserve power to the greatest extent possible. Sound processing that conserves power is effective because A/D converters, microphone amplifiers and the sound processing unit consume a certain amount of power even when they are sensing no sound.

We have proposed a low-power VAD hardware implementation using a single microphone [6]. This custom hardware uses a zero-crossing algorithm for the VAD. Fig. 3 portrays the zero-crossing algorithm, as implemented on an FPGA in the ubiquitous sensor node as well.

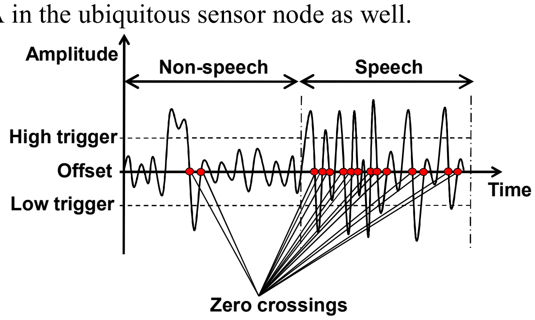

Fig. 3. Zero-crossing algorithm. The offset line shows the direct current (DC) component.

The zero crossing is the first intersection between an input signal and an offset line after the signal crosses a trigger line: the high trigger line or the low trigger line. Between a speech signal and non-speech signal, the appearance ratios of this zero crossing differ. The zero-crossing VAD detects this difference and outputs the beginning point and the end point of a speech segment.

· *Sound Source Localization (SSL)*

For SSL, we propose a hierarchical localization method. We divide localization into two layers: 1) relative direction estimation within a node, and 2) absolute location estimation by exchanging results through the network.

The MUSIC algorithm [7] is chosen for node layer estimation because microphones on the node are limited to 16; the MUSIC algorithm can achieve higher resolution with fewer microphones. To find a relative direction, the sound source probability for $P(\theta, \phi)$ is calculated for each node. Once the relative localization data are obtained, they are sent to a neighboring node to proceed to the next step.

We will localize the absolute sound source location in the network layer. A brief description of this method is presented in Fig. 4 with a three-dimensional coordinate of the sound source.

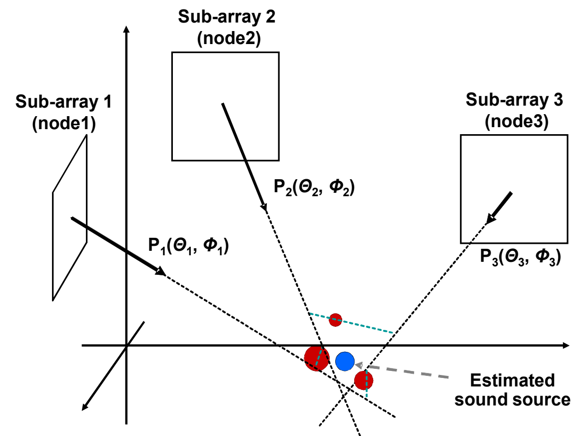

Fig. 4. Three-dimensional sound source localization (SSL).

Each node calculates $P(\theta, \phi)$ using the MUSIC algorithm and finds the maximum of $P(\theta, \phi)$ and corresponding θ and ϕ. Then, the localization data (a set of θs and ϕs) are exchanged, and we adopt the shortest line segment that connects two lines. We infer a point that divides the shortest line segment by the ratios of $P(\theta, \phi)$s as an intersection. The sound source is localized by calculating the center of gravity of these weighting intersections.

Error! Not a valid link.

Fig. 5. Sound source localization (SSL) accuracy.

We verified the hierarchical localization through simulation, assuming a result of relative direction estimation has a variation on every node. Fig. 5 shows the localization accuracy. The localization error is as much as 10 cm and becomes smaller when the arrays are numerous and the direction estimation is precise. The results show that minimization of the direction error increases the localization accuracy effectively.

· *Sound Source Separation (SSS)*

Geometric techniques with position information and statistical techniques using no position information are two major SSS methods. The proposed system uses a basic geometric method, delay-and-sum beamforming [8], because the node positions are known. This method produces less distortion than statistical techniques; moreover, it requires few computations (only interpolation and summation). For distributed processing for sound source separation, it can be applied easily because it is based on the interpolation and summation (Fig.6).

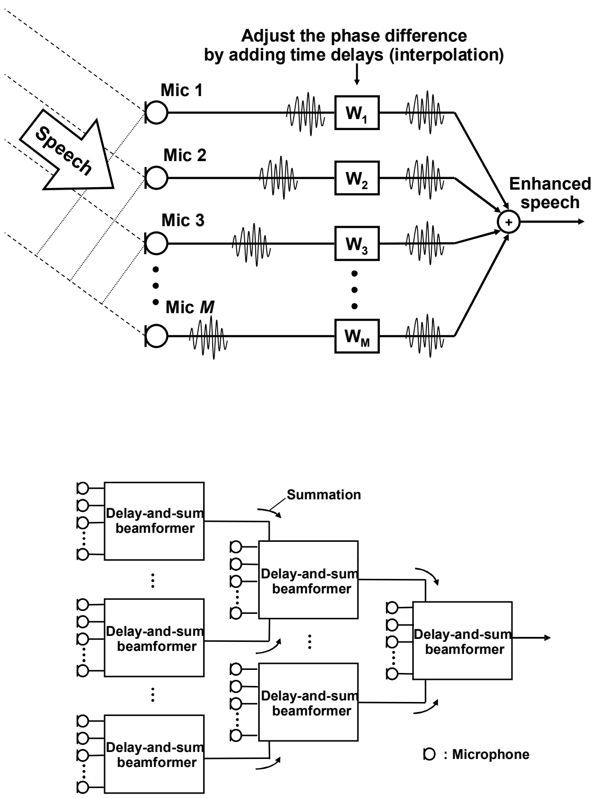

Fig. 6. Delay-and-sum beamforming with a node and among nodes.

4. ARCHITECTURE AND IMPLEMENTATION

Regarding practical design, we implemented the intelligent ubiquitous sensor node on an FPGA board (SZ410, Suzaku; Atmark Techno Inc.) and microphones (ECM-C10; Sony Corp.). Fig. 7 shows prototype system photographs.

Fig. 8 presents the breakdown of the processing workload in the VAD, SSL and SSS. We adopt parallel-processing DSP architecture for the SSL because it deals with fixed-point matrix operation such as summation, subtraction and multiplication; the SSL dominates the workload in a node. For the SSS, we merely prepare an integer multiplier (interpolator) and adder. Fig. 9 presents the system architecture.

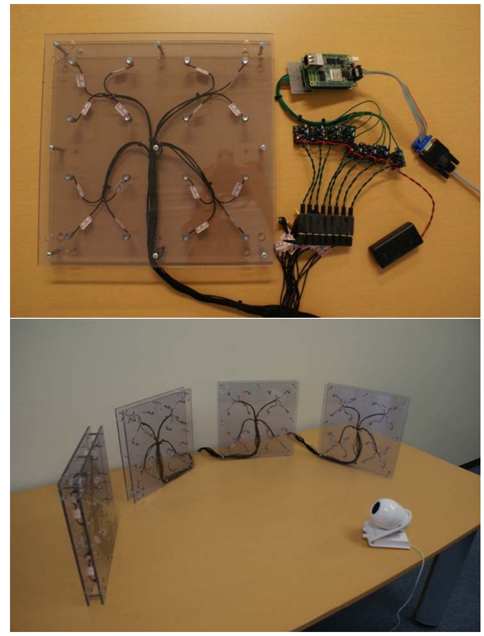

Fig. 7. System photographs: intelligent ubiquitous sensor node and a microphone array comprising sub-arrays.

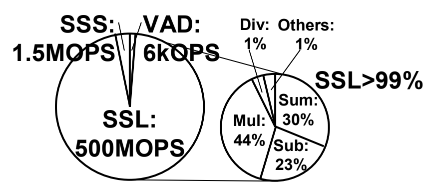

Fig. 8. Workload breakdown.

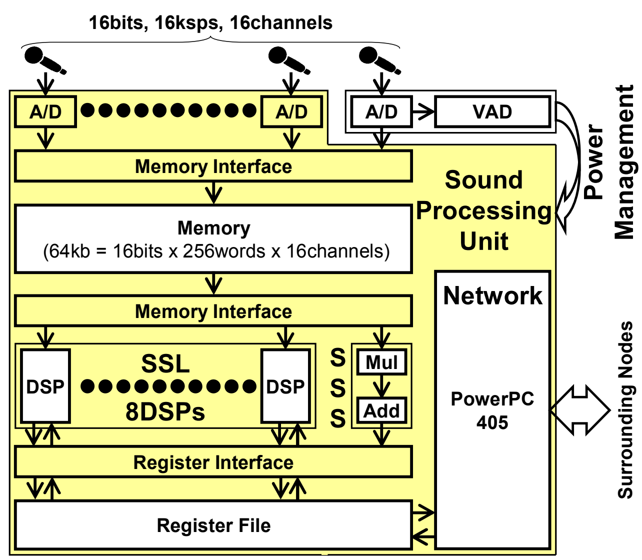

Fig. 9. System architecture.

By separating the low-power VAD module from the sound processing unit, it can manage the power-hungry sound processing unit using clock gating. The power management is particularly important in future LSI design for a sensor node. Note that a single microphone is sufficient to detect a signal: the other 15 microphones can be turned off. Furthermore, not all VAD modules in all nodes need operate: the system activates only some.

Fig. 10 portrays the block diagram of the VAD module. All calculations are integrated using integer arithmetic. In our simple VAD algorithm, the sampling frequency can be reduced to 2 kHz and the number of bits per sample can be set to 10 bits. These values are sufficient to detect human speech, in which case only 2.1 mW is dissipated on the FPGA. The slice numbers of flip flops and 4-input LUTs are, respectively, 1,015 and 3,831. Fig. 11 shows input waveforms and the VAD results for an SNR of 0 dB, where the accuracy of the VAD is 96%.

Error! Not a valid link.

Fig. 10. Block diagram of VAD.

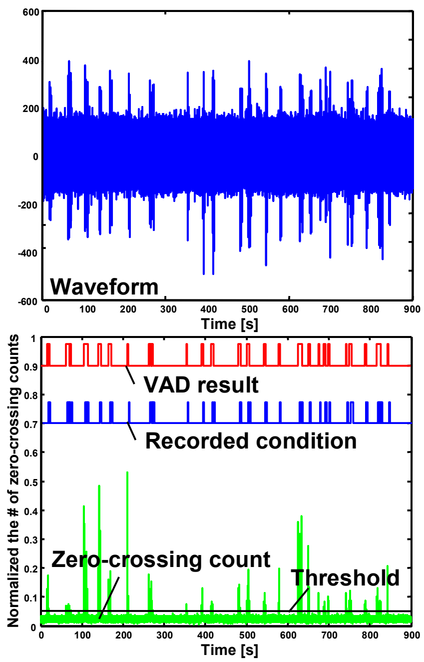

Fig. 11. Waveform and VAD result at SNR = 0 dB.

Fig. 12 shows that SNR improvement of 7.75 dB was gained with 112 microphones (seven sub-arrays). We anticipate 15 dB or greater improvement using several hundreds of microphones or several tens of sub-arrays.

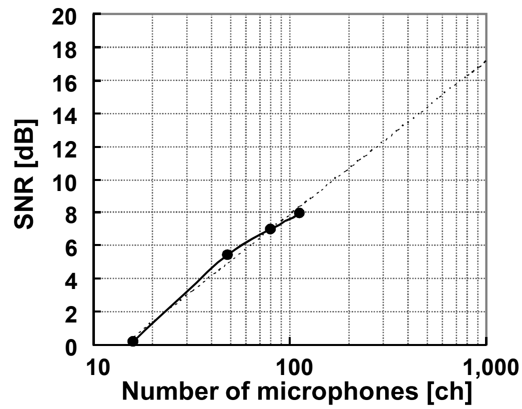

Fig. 12. SNR vs. number of microphones.

5. CONCLUSION

This proposed microphone array network realizes ubiquitous sound acquisition. A microphone array network comprising 16-microphone sub-arrays performed the following three operations within a node and with a network: 1) low-power voice activity detection (VAD) to activate the entire node, 2) sound source localization (SSL) to find sound sources, and 3) sound source separation (SSS) to enhance the sound. Low-power VAD was implemented to manage the nodes' power consumption. Thereby, the system achieves low power when speech is not active. The VAD module dissipates only 2.1 mW on an FPGA. The SSL is processed with the distributed nodes. The experimental result of the SSS demonstrated an SNR improvement of 7.75 dB using 112 microphones. The system will achieve an SNR of 15 dB if the entire microphone network has more than several hundred microphones.

REFERENCES

[1] E. Weinstein, K. Steele, A. Agarwal, and J. Glass, "Loud: A 1020-node modular microphone array and beamformer for intelligent computing spaces," *MIT/LCS Technical Memo*, MIT-LCS-TM-642, 2004.

[2] Yuuki Tamai, Satoshi Kagami, Hiroshi Mizoguchi, Yutaka Amemiya, Koichi Nagashima, and Tachio Takano, "Real-Time 2 Dimensional Sound Source Localization by 128-Channel Huge Microphone Array," *IEEE RO-MAN*, pp. 65-70, Sep. 2004.

[3] Y. Tamai, S. Kagami, H. Mizoguchi, K. Sakaya, K. Nagashima, and T. Takano, "Circular microphone array for meeting system," *IEEE Sensors*, vol. 2, pp. 1100-1105, Oct. 2003.

[4] M. Brandstein and D. Ward, "Microphone Arrays: Signal Processing Techniques and Applications," *Springer*, 2001.

[5] H. F. Silverman, W. R. Patterson III, and J. L. Flanagan, "The Huge Microphone Array," *IEEE Concurrency*, vol. 6, pp. 36-46, Oct.-Dec. 1998.

[6] H. Noguchi, T. Takagi, M. Yoshimoto, and H. Kawaguchi, "An Ultra-Low-Power VAD Hardware Implementation for Intelligent Ubiquitous Sensor Networks," *IEEE SiPS*, pp. 214-219, Oct. 2009.

[7] F. Asano, H. Asoh, and T. Matsui, "Sound source localization and signal separation for office robot (Jijo-2)," *IEEE MFI*, pp. 243-248, Aug. 1999.

[8] J. Benesty, M. M. Sondhi, and Y. Huang, "Handbook of Speech Processing," *Springer*, 2007.