April 19-24, 2009

IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings

TAIPEI, TAIWAN

ICASSP 2009 TAIPEI

Signals over the Horizon

IEEE
Celebrating 125 Years
of Engineering the Future

IEEE Signal Processing Society

2009 IEEE International Conference on Acoustics, Speech, and Signal Processing

PROCEEDINGS

April 19—24, 2009
Taipei International Convention Center
Taipei, Taiwan

Sponsored by

The Institute of Electrical and Electronics Engineers
Signal Processing Society

IEEE Catalog Number: CFP09ICA
ISBN: 978-1-4244-2354-5
ISSN: 1520-6149

IEEE Catalog Number: CFP09ICA
ISBN: 978-1-4244-2354-5
ISSN: 1520-6149

Assembled by Conference Management Services, Inc.

IEEE

A PORTABLE USB-BASED MIROPHONE ARRAY DEVICE FOR ROBUST SPEECH RECOGNITION

*Qi (Peter) Li, Manli Zhu, and Wei Li**

Li Creative Technologies, Inc. (LcT)
30 A Vreeland Road, Suite 130, Florham Park, NJ 07932, USA
E-mail: li,manlizhu@licreativetech.com; *iewil2000@yahoo.com

ABSTRACT

We present a USB-based, highly directional, and portable microphone array device that delivers a crisp, clear and noise-reduced speech signal. This device consists of four linearly distributed microphone sensors and a filter-and-sum beamformer designed using broadband beam-forming algorithm. The device has a narrow acoustic beam pattern and identical frequency responses for almost all speech bands. In addition to beamforming, an adaptive noise reduction algorithm is used to further reduce the background noise. By utilizing both the spatial and temporal information, the SNR of speech signals is improved and speech recognition performance in noisy environments is significantly improved as reported in our experiments.

Index Terms **—** Microphone array, beamforming, noise reduction, robust speech recognition.

1. INTRODUCTION

A microphone array consists of multiple microphone sensors located at different positions. It can be used for both sound source location [1] and speech enhancement by processing the signals from each individual sensors [2][3]. While most of the current speech processing software can only use the temporal information, the designed device utilizes both the spatial and temporal information; thus, significantly improving speech recognition performance and robustness.

One of the major challenges in applying a microphone array in speech recognition is that speech is a wideband signal. The traditional narrowband beamforming techniques are not appropriate anymore [4]. The problem has been addressed by the spatial Fourier transform of a continuous aperture [5] and the joint optimization of the spatial and frequency response [6]. In these approaches, to keep the constant response over the wide frequency range, the array size is usually large; thus most of the prototypes or products using microphone arrays on the market are quite large and cannot be used as a portable device [7]. The large size of the array prevents the array products from broad applications, such as handheld devices, wireless handsets, and PDA. Another challenge of a microphone array compared to single close-talking microphones is the decreased performance in speech recognition because of the variation of their frequency responses [8]. These problems need to be addressed by improving the array design algorithms, in order to develop a portable and high performance device.

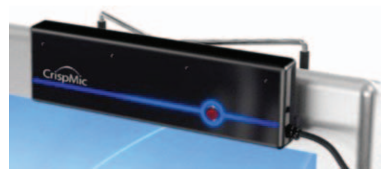

Figure 1. The 4-sensor microphone array named CrispMic™ clipped on a laptop

In this paper, we present a portable microphone array device, named CrispMic™, where both the size and performance have been properly addressed. As shown in Figure 1, the size of the microphone array is only 9.6 x 2.6 x 1.3 cm which is less than the half of the size of a similar microphone array on the market. Also, it has a constant frequency response in a wide range of speech bands as shown in Figure 4. The experimental results in this paper show that the array can improve speech recognition performances significantly, especially in noisy environments.

2. SYSTEM DESCRIPTION

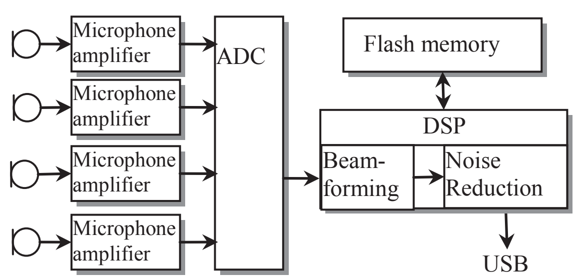

Figure 2. Illustration of the hardware structure

Figure 2 shows the structure of the device. It consists of four identical omni-directional microphones, an audio codec chip with an ADC and pre-amplifiers, a DSP (digital signal processor) chip, a flash memory and a USB interface. The acoustic signal is picked up by four microphone components arranged as a linear array with 20 mm intervals. The audio codec provides an adjustable gain, and converts the four channels of analog signals into digital signals for the DSP. The beamforming algorithm combines the 4 channels of speech into one channel and then a noise reduction algorithm is applied to further reduce the background noise. The processed clean speech signals are then transmitted to a laptop or PC through the USB interface.

The flash memory stores the software code for the DSP chip. Once the system boots up, the DSP chip reads the code from the flash memory into the internal memory and starts to execute the code. The device is powered through the USB. It is a plug-and-play device and can be used without installing any software. Since both the analogue and digital circuits are implemented in a small PCB, the circuit board was especially designed to avoid the noise interferences.

2. BROADBAND BEAMFORMING

Due to the special requirements in size and performance, we developed a robust, far-field broadband beamforming algorithm and implemented it in the DSP chip. The beamformer has a constant response in the speech frequency bands between 300-4000Hz. In theory, it can significantly improve spatial SNR of the speech signals without distortions in different frequency bands.

The linear microphone array configuration is shown in Figure 3. It is comprised of four equally spaced microphone sensors, where d_i is the distance between the i^{th} microphone and the center of the array. The output y of the array is the filter-and-sum of the four microphone outputs, $y = \sum_{n=1}^{4} w_n^T x_n$.

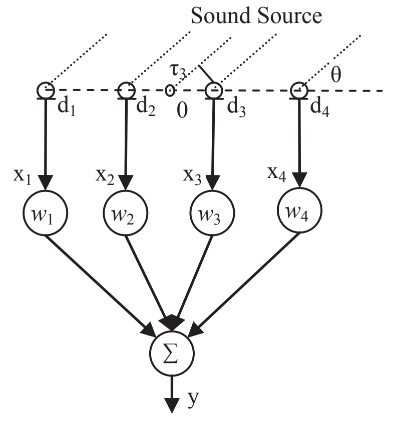

Figure 3. The configuration of linear microphone array.

The spatial directivity pattern $H(\omega,\theta)$ for the sound source form angle θ with normalized frequency ω is defined as [2]:

$$H(\omega,\theta) = \frac{Y(\omega,\theta)}{\overline{X}(\omega,\theta)} = \frac{\sum_{n=1}^{4} W_n(\omega) X_n(\omega,\theta)}{\overline{X}(\omega,\theta)} \quad (1)$$

where $\overline{X}$ is the signal received at the center of the array and W is the frequency response of the real-valued FIR filter w. If the sound source is far enough from the array, the difference between the signal received by the n^{th} microphone x_n and the center of the array is a pure delay. We use $\tau_n = f_s d_n \cos\theta / c$ to measure the delay by the number of samples, where f_s is the sampling frequency, c is sound speed, and $X_n(\omega,\tau) = \overline{X}(\omega,\theta) e^{-j\omega\tau_n}$ is the microphone signal. The spatial directivity pattern H can be re-written as:

$$H(\omega,\theta) = \sum_{n=1}^{4} W_n(\omega) e^{-j\omega\tau_n(\theta)} = \mathbf{w}^T \mathbf{g}(\omega,\theta) \quad (2)$$

where $\mathbf{w}^T = [\boldsymbol{w}_1^T, \boldsymbol{w}_2^T, \boldsymbol{w}_3^T, \boldsymbol{w}_4^T]$ and $\mathbf{g}(\omega,\theta)$ is the steering vector.

Let the desired spatial directivity pattern equal 1 in the pass band and 0 in the stop band. The cost function is then defined as:

$$J(w) = \int_{\Omega_p}\int_{\Theta_p} |H(\omega,\theta) - 1|^2 \, d\omega \, d\theta + \alpha \int_{\Omega_s}\int_{\Theta_s} |H(\omega,\theta)|^2 \, d\omega \, d\theta \quad (3)$$

Let $\partial J / \partial w = 0$. We can then obtain the best parameter set w.

A Computer simulation was conducted to verify the performance of our designed beamformer with the following parameters: The distance between microphones is 0.02m, the sampling frequency f_s = 48k, and FIR filter taper length L=128. When the pass-band (Θ_p, Ω_p) = {300-4000Hz, 70°-110°}, the designed spatial directivity pattern is 1. When the stop-band (Θ_s, Ω_s) = {300~4000Hz, 0°~60°+120°~180°}, the designed spatial directivity pattern is 0.

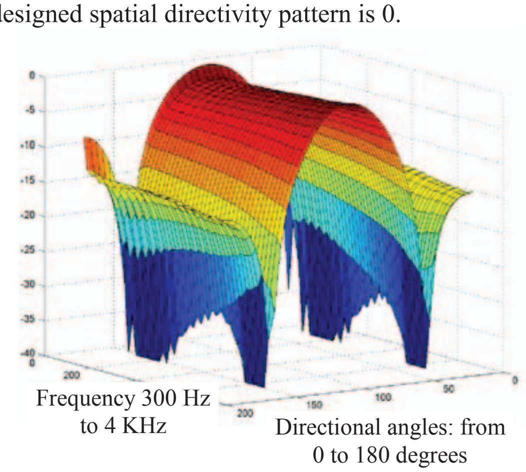

Figure 4. Directivity pattern of the designed microphone array: the frequency bands from 300 Hz to 4 kHz of the sound from the front of the microphone array are enhanced and the sound from other directions are reduced by about 15 dB.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase
Smarter legal research.