Wearable Microphone Array as User Interface

Yong Xu , Mingjiang Yang, Yanxin Yan, Jianfeng Chen

Institute for Infocomm Research, Singapore

xuyong@i2r.a-star.edu.sg

Abstract

We are at present enabled with machine-empowered technologies. The future is certainly looking towards human-empowered technologies, which should enable mobile user with natural wearable devices along with natural-like user interfaces. This paper presents proof of concept of wearable microphone array system that could be embedded in textiles for hands-free, obtrusion-free and hassle-free operation interface with body-worn computer, which is replaced with a speech recognition engine with a PC for testing. One mode of microphone array is for user-only speech interface with the computer and another mode is to collect others voice and steer it by voice control to get maximum sensitivity. The details of microphone array systems are beyond the scope this short paper. The concepts and the preliminary promising results are presented here to register our thoughts with 'wearable computers/ user interface' community.

Keywords: microphone array, wearable, speech interface

1. Introduction

We are at present living with machine-empowered technologies. The future is certainly looking towards human-empowered technologies, which should enable mobile user with natural wearable devices along with natural-like interfaces. Significant progress has been reported on wearable computers in recent years with low form factor [1]-[4]. However, wearable computers have not picked the expected market segment for the reasons that they are still obtrusive, heavy, not hands-free and not so user friendly. According to Venture Development Corporation (VDC) and other visionaries, the true potential for wearable computing would pick-up if improvements are made in consumer-based products, including commercially viable 'smart fabric' technology, particularly a new class of wearable systems made of fabric.

The processing hardware and software could be distributed into other natural wearable. Then the user looks natural, i.e. no different from public, which in turn results in enormous impact of commercial applications.

(a) **(b)**

Figure 1: Close-talking microphone

Speech driven user interface to outside world (i.e. wearable computer embedded in e-textiles) is an important subsystem. Generally, to provide high speech quality, close-talking microphone is the best choice to interface voice commands to the wearable computer or to a machine. However, one form of close-talking microphone needs a hand to hold the device near the lips as shown in the figure 1(a) and hence it is not hands-free and also not restriction-free from the point of movement of the head. Another form of close-talking microphone appears as a headgear as shown in the figure 1(b). This is widely used as 'hands-free and restriction-free' device. However, it is an obtrusive arrangement and is not in tune with future developments.

Wearable Microphone Array (WMA) is an important system to interface context aware speech processing to the wearable computer. There are methods but limited performance to capture the user's voice in presence of high ambient noise (commands should be recognized by machines) and also to listen to other speakers.

Other microphone array systems are available for Human Computer Interactions (HCI) [5][6] but are not suitable for mobile users.

The need for microphone arrays instead of a collar-microphone and our proposed concept are presented in section 2, include the structure description of two modes of microphone arrays. Some experimental results are presented in section 3 and the conclusions and the future work are presented in section 4.

2. Proposed Wearable Microphone Array

In order to realize 'no impact on appearance' and 'no impact on comfort', wearable microphone array (WMA) and its related modules should be strategically designed. On the one hand, the structure should be in accordance with people's living habit and be convenient to be integrated or embedded inside clothes. On the other hand, the arrangement should be in favor of noise reduction and interference attenuation. Bearing these requirements in mind, an initial scheme by putting multiple sensors in a tie-shape textile is proposed as follows.

2.1 Structure

As shown in figure 2, microphone sensors will be embedded into the textile and therefore be invisible.

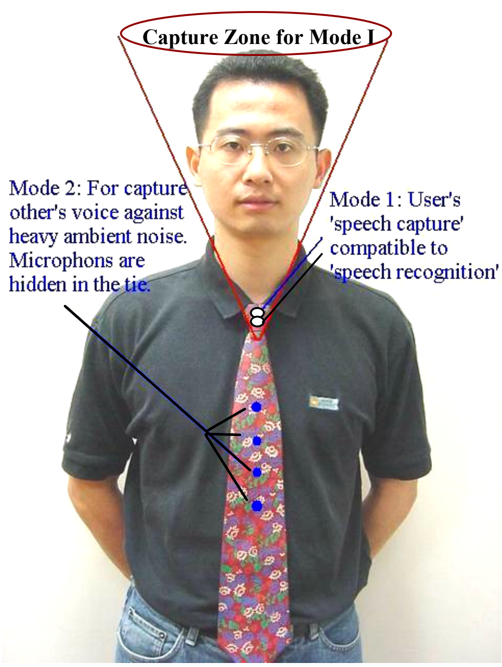


Figure 2: Structure of proposed wearable microphone array embedded in a tie

The whole structure is composed of two parts corresponding to the two functions. The purpose of the part one (mode-I), as shown in the knot of the tie, is to pick up the voice from the user himself and to effectively reduce the noise around him. The part one can also be embedded into a shirt collar. On the other hand, the part two (mode-II) will be used to capture the voice from other speakers and it is hidden in the main body of the tie. It can be alternatively embedded into the load-bearing column of a shirt in form of buttons.

2.2 Mode-I

As mentioned above, mode-I is used only for capturing the user's voice. To ensure that the user could be able to command the computer freely and precisely, the captured voice must be clear enough and be recognizable to speech recognition module in his wearable computer. Therefore the system should be able to effectively reduce the surrounding background noise and interference and at the same time retain the original speech undistorted.

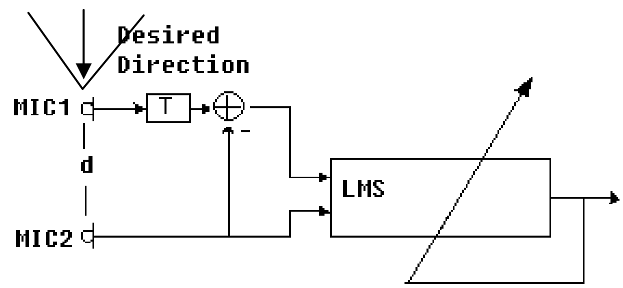


Figure 3: Diagram of adaptive microphone array structure

As shown in figure 3, mode-I is realized by using two far field miniature omni-directional microphones arranged in a small space. The two microphones are located in an end-fire orientation. T represents the sound propagation delay from MIC 1 to MIC 2 and d denotes the distance between the two microphones.

LMS module has two inputs: reference noise signal and MIC 2's signal. The reference noise signal, which theoretically contains only undesired signal, is obtained from the output of an intermediate first-order differential microphone constructed from MIC1 and MIC2. The reference signal is then adaptively subtracted from the output of MIC 2, which is a mixture of desired and undesired signal, by implementing the LMS algorithm as in [7]. The updating of the weighing coefficients in the adaptive algorithm is based on the coherence function [8].

As a result, the array functions as a virtual directional microphone with cone-shape directionality towards user's head, as illustrated in figure 2. It provides high spatial attenuation to the sound sources coming outside the cone.

2.3 Mode-II

Mode-II is designed to listen to the voice of other speakers/machines. We start with a linear array and a classical adaptive beamforming algorithm working in frequency domain is adopted and implemented, [9] which we will not go through in detail in this paper. Detailed structure and algorithm can be found in [9] and [10].

3. Experimental Results

A series of experiments have been carried out to demonstrate the performance of our proposed scheme and establish the need of microphone array system in wearable computer system. figure 4 shows the basic test set up. Two loud speakers are around the user and played noise and interfere voice when the user speaks.

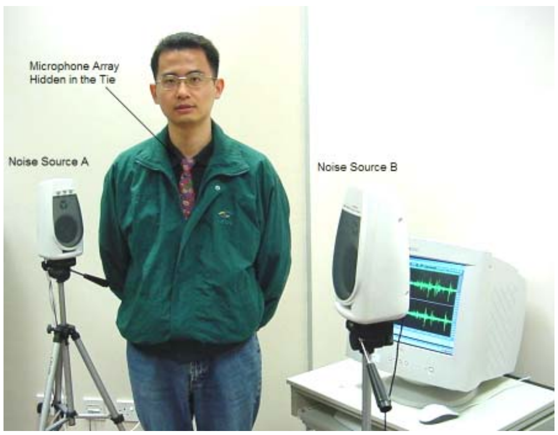


Figure 4: Test setup for mode-I

In the first experiment, nine menu commands (File, Edit, View, Insert, Format, Tools, Table, Window and Help) in Microsoft Office XP voice command system was used as a basic speech recognition test platform. We used different types of close-talking microphone and compared the results with that using mode-I in different situations which include (a) Noise free; (b) white noise plus interfere voice (SNR 10dB). The summaries of speech recognition correct rates are recorded.

Condition	Ratio
Close-talking Mic; No Noise	100%
Close-talking Mic; with Noise and interfere	71%
Mode-I; No Noise	91%
Mode-I; White noise and interfere	81%

Table 1: Speech Recognition ratio in different situations

Similar tests were carried out for a number of users and the average results are given in Table 1. The performance of close talking microphone is 100% in noise free conditions where as the performance of mode-I is only 91%. The performance loss in the later is due to beam formation and associated processing. However, the performance of mode-I is 81% in noisy conditions much better compared to close talking microphone 71%.

The waveforms captured in noisy condition and processed outputs are shown in figure 5. In the test, the noise includes white noise and cross-talking noise at the same time. The result below shows that the user's voice can be extracted from the input with very low SNR (no more than 5 dB) by the system.

Tests are also conducted as the user changed the direction of his face while talking. The results are very satisfactory attenuating the noise from even nearby sources. There are 4 groups in the figure 6. Each group includes two waveforms. The top one is the input of the system that includes the user's voice and heavy ambient noise. Another one is the output of the system. It is shown clearly by figure 6 that the ambient noise has been restrained no matter which direction the user faces.

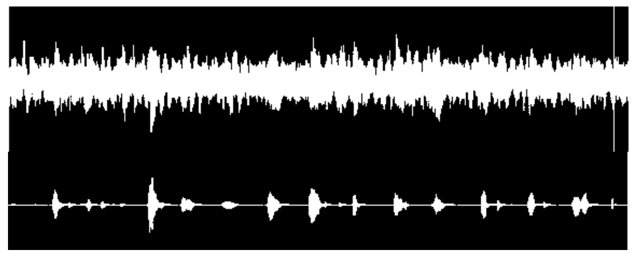
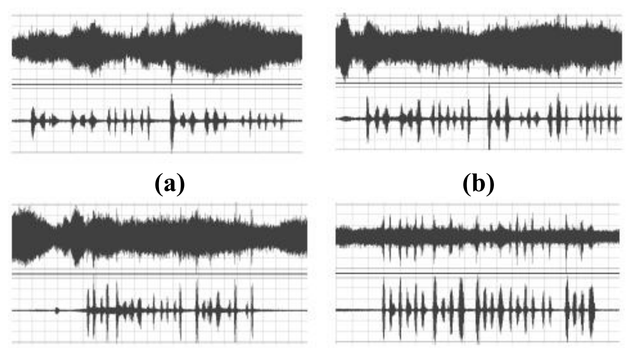


Figure 5: Mode-I waveforms

Our analysis indicate that the performance of close talking microphone is not at all satisfactory in noisy conditions where as our conceptual mode-I is much better. We also conducted spectral analysis of our mode-I why certain words are not recognized properly. We found that the frequency response of our mode-I starts from 350Hz which need to be improved. It is a trade of between canceling noise versus improvement in frequency response.



(a) **(b)** **(c)** **(d)**

Figure 6: Mode-I test results:
(a) face right; (b) face left; (c) face up; (d) face down

Similarly we have conducted experiments on Mode-II. Figure 7 shows the waveform that before and after the processing. The noise was reduced and the speech quality was increased by the system.

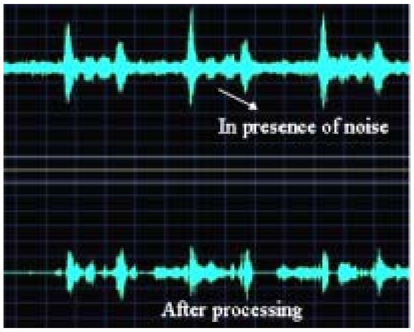


Figure 7: Mode-II results

4. Conclusion

We envisage user interface technologies to enhance the future mobile users needs with hands-free, obtrusion-free and hassle-free functions. And also the user, while wearing the user interface devices should look no different from the public. In that direction, we developed new algorithms and configurations on 'wearable microphone array systems" which can be embedded in textiles eventually. I our experiments, we have stitched the required microphones on to the tie physically but all the wires carrying the signals are brought to HW/PC for further processing. We are trying to collaborate with Georgia Institute of Technology, USA to embed the microphones as well as the processors within the tie. The preliminary test on mode-I and mode-II are very satisfactory. Mode-I is particularly important for voice driven computer interface.

The focus of this paper is to develop 'proof of concept' leading to embedding them into textiles. However, we need to carry out rigorous experiments with different types of noise and quantify the performance figures.

We would like to improve the performance of mode-I to 99% in no noise condition and at least to 90% in presence of noise (SNR=10dB). We also would like to see the fusion effect of mode-I and mode-II. Our far future work is to 'model the microphone array system' as how the performance degrade when the inter-distance between microphones vary as a result of textile flexibility, particularly mode-II.

5. References

[1] The PC goes ready-to-wear: Cover story IEEE spectrum, Oct 2000

[2] Xybernaut's Poma™, a wearable computer, http://www.fashionwindows.com/fashion/2002/wearable_comp.asp

[3] Sumit Basu, Steve Schwartz, and Alex Pentland. "Wearable Phased Arrays for Sound Localization and Enhancement." In *Proceedings of the IEEE International Symposium on Wearable Computing (ISWC '00)*. Atlanta, Georgia. October, 2000. pp. 103-110

[4] Nitin Sawwhney and Chris Schmandt, "Nomadic Radio: Speech & Audio Interaction for Contextual Messaging in Nomadic Environments", special issue of ACM Transactions on Computer Human Interaction (ToCHI) on the topic of Human Computer Interaction and Mobile Systems. June'2000

[5] Thomas M. Sullivan, " Multi-Microphone Correlation-Based Processing for Robust Automatic Speech Recognition", PhD thesis 1996

[6] Xu Yong, Competitive intelligence/research intelligence/market intelligence on microphone arrays", Internal report, Institute for Infocomm Research, Singapore

[7] Hoshuyama, A. Sugiyama, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," IEEE Trans. on Signal Processing, vol. 47, no. 10, pp. 2677-2684, 1999

[8] J.J. Shynk, "Frequency-domain and multirate adaptive filtering", IEEE Signal Processing Magazine, pp. 14-37, 1992

[9] W. H. Neo, B. Farhang-Boroujeny, "Robust microphone arrays using subband adaptive filters," "IEE Proc.-Vis. Image Signal Process. vol. 149, no. 1, pp.17-25, 2002

[10] Hoshuyama, A. Sugiyama, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," IEEE Trans. on Signal Processing, vol. 47, no. 10, pp. 2677-2684, 1999