5.1-3

Hands-Free Voice Communication Platform Integrated With TV

Istvan PAPP, Zoran ŠARIĆ, Saša VUKOSAVLJEV, Nikola TESLIĆ and Miodrag TEMERINAC
Faculty of Technical Sciences, Novi Sad, Serbia

***Abstract*-- This paper presents a system for full-duplex hands free voice communication integrated with TV technology. The system provides comfort conversation by utilization of microphone array and advanced voice processing algorithms, even with simultaneous TV usage. As communication channel GSM or VoIP can be used.**

I. INTRODUCTION

The modern means of digital voice communication are more and more present in the consumer market. They become more affordable every day, and plays significant role in both business and private life. In order to provide comfort communication, hands-free systems are widely used.

Usage of such systems implies unspecified talker position in the acoustic environment, with variable distances from system's microphones and loudspeakers. Hands-free speech communication in such unspecified conditions involves a number of technical problems, which must be solved in order to preserve good quality of speech communication.

This paper describes a system that provides full-duplex hands-free communication in very complex acoustic environment. The typical use case scenario is depicted on Fig. 1. The developed system makes possible placing and accepting calls from remote parties. It also provides contacts management through the intuitive graphical interface rendered on the TV screen. The system connects to a gateway via Bluetooth. As gateway, either GSM phone or PC can be used. In latter case, on the PC a VoIP application is active.

Fig. 1. Hands-free voice communication platform integrated with TV

An innovative feature is that the system can be used while watching TV broadcast. The voice of the remote party is played back on the TV loudspeakers, mixed intelligently with the TV broadcast sound. The remote party will receive a high quality voice of the active speaker, with eliminated acoustic disturbances like echo, noise or TV show sound. During the whole session the communication is full-duplex.

II. PLATFORM OVERVIEW

In order to efficiently fight with acoustic disturbances, the system uses a microphone array of 5 elements. The signals of the microphones are processed by a DSP in real time. The developed algorithms suppress the disturbances, leaving only the desired speech. The improved voice is transmitted to the gateway (GSM phone or PC) via Bluetooth, and then to the remote. The DSP and the connectivity module are located on an add-on module (see Fig. 2), which can be easily interfaced with the host. The tasks of the host are to route audio channels appropriately, to control the connectivity module and to provide an interface to the user.

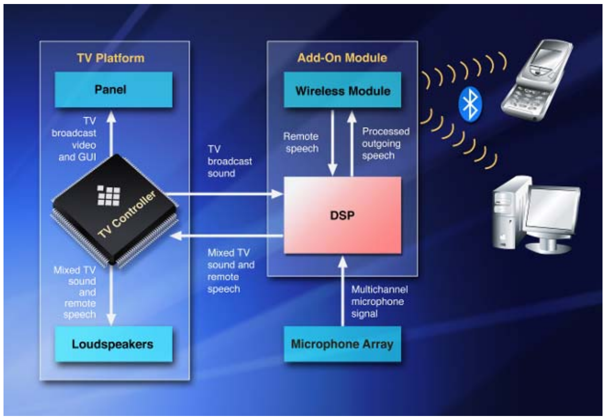

Fig. 2. Integration of phone functionality with TV

III. AUDIO PROCESSING ALGORITHMS

Using TV as a host induced numerous problems due to the fact that the loudspeaker and microphone array are closer to each other than the distance between the user and the TV unit. This leads to a very complex acoustic environment, as shown in Fig. 3. In such set-up there are several types of disturbances, such as strong echo coming from the loudspeakers, high reverberation time due to the room dimensions, presence of diffuse and spatially allocated non-stationary noise sources, as well as a low SNR.

The microphone array is coupled with set of advanced voice processing algorithms. The multichannel acoustic echo canceller (AEC) is an adaptive NLMS structure based on FIR filter [1]. It cancels the sound played back on loudspeakers in the microphone signals. The filter is long enough to handle room reverberation time up to 300 ms. The adaptation control module provides fast filter adaptation to the changing environment. The adaptation is controlled by a sophisticated double talk detector (DTD) that provides soft indication of near end speaker activity. Loudspeakers signals are recorded by own analog/digital converters to makes system robust against packet loss when VoIP protocol is used.

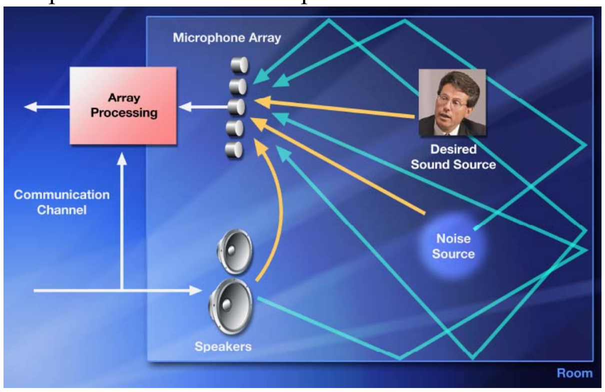

Fig. 3. Disturbances in acoustic environment

The non-adaptive superdirective beamformer extracts the sound coming from desired direction, matching the direction of the active speaker [2]. This information is provided by the direction of arrival (DOA) block, which finds the dominant speaker in the horizontal plane [3]. DOA is based on generalized cross-correlation approach and phase transformation, combined with voice activity detector. By the spatial filtering, the effects of reverberation are significantly reduced and the spatially arranged noise sources are suppressed. The noise reduction deals with the stationary ambient noise. It is based on the approach described in [4]. The automated gain control block ensures that the level of the output voice is of constant power. It is a novel dynamic range compressor, which utilizes the information about spatial energy distribution to identify near end speech segments that should be amplified or attenuated.

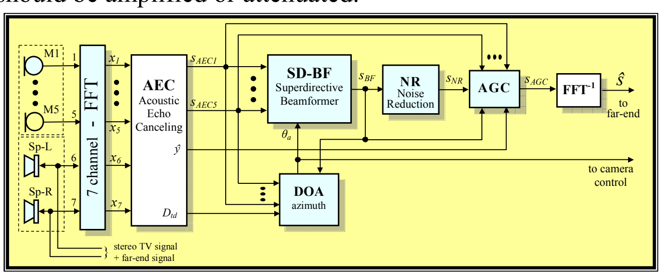

Fig. 4 –Set of voice processing algorithms

The algorithms are implemented as a dedicated hardware module using DSP[1] and optimized for real-time performance at sample rate of 8 kHz and microphone array of 5 elements.

IV. RESULTS

The algorithm development for audio processing requires systematic approach and development of the automatic procedures for signal quality measurement. With objective metrics like ERLE, SNRE, PESQ [5] systematic monitoring of the audio signal quality is ensured. As testing environment a living room-like ambient was used, with reverberation time of 300 ms, and spatially allocated noise sources. The communication endpoints were 3.5 meters from the microphone array, while the loudspeaker were located 0.5 meters beneath the array. The objective measures ERLE and SNRE resulted with 30 dB suppression of echo and noise, while PESQ was 2.8 for both single and double talk conditions, compared to 1.3 without algorithms applied.

The relatively small difference in PESQ score caused significantly higher subjective quality impression. This is verified in the nonsense syllables recognition test, with running TV broadcast in the background. The percentage of correctly recognized nonsense syllables in the tests was 24% without speech improvement (Series_A), while with advanced audio processing (Series_B) the results increased to 49% (Fig.5).

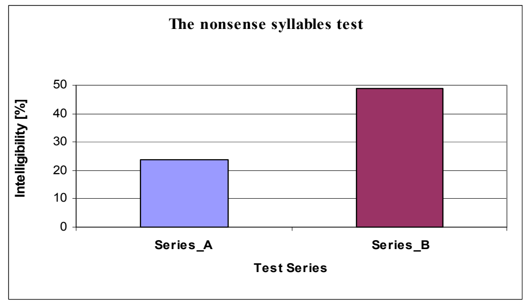

Fig. 5 – The nonsense syllables test results

V. CONCLUSIONS

The developed system provides high quality full-duplex voice communication in hands-free mode suitable for consumer electronics applications. The targeted use case is the environment like living room or office. The developed audio subsystem is integrated with TV. This lead to a hands-free communication terminal with advanced features like simultaneous TV usage, full-duplex operation and high voice quality at affordable price. It makes possible comfort conversation using any of the communication technologies (GSM or VoIP). The developed technology can be used in systems like car hands-free kit, teleconferencing system, as well as in voice based human-machine interfaces [6].

REFERENCES

[1] S. Haykin, B. Widrow, "Least-Mean-Square adaptive Filters", Wiley, 2003.

[2] I. Papp, Z. Saric, S. Jovicic, N. Teslic, "Adaptive microphone array for unknown desired speaker's transfer function", Journal of Acoustic Society of America, Express Letters, pp. 44-49, July 2007.

[3] D. Kukolj, M. Janev, I. Papp, N. Teslić and S. Vukobrat, "Speaker Localization under Echoic Conditions Applied to Service Robots", EUROCON 2005, Beograd.

[4] Israel Cohen, "Noise Estimation by Minima Controlled Recursive averaging for Robust Speech Enhancement", IEEE Signal Processing Letters, Vol. 9, No. 1, January 2002, pp. 12-15.

[5] ITU-T, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", International Telecommunications Union, 2001.

[6] I. Papp, D. Kukolj, Z. Marčeta, V. Đurković, M. Janev, M. Popović, N. Teslić, "Remotely Controlled Semi-Autonomous Robot with Multimedia Abilities", ICCA 2005, Budapest, 2005.