[1]A Compressive Sensing-based Speech Signal Processing System for Wearable Computing Device in IPTV Environment

Kuei-Hong Lin, Cheng-Hsun Lin, Kuo-Huang Chung and Kai-Shun Lin

Abstract A design of compressive sensing (CS)-based speech signal processing (SSP) system for wearable computing device (WCD) in internet protocol television (IPTV) environment is proposed in this study. In WCD, the voice control would be an important function since the WCD is too small to provide typing device. Hence, the quality of the recorded user's comments will determine the user experience (UX) of the interactions between the user and the WCD. For building a user-friendly entertainment environment of the IPTV, the proposed SSP system integrates the WCD to be the medium of the voice control, that is, the voice recorder of the WCD would record the background voice for further analyzing continuously, and it would waste lots of the bandwidth and power of the WCD to transmit the data to the cloud. Therefore, the proposed method is applied to compress the recorded voice for saving the transmitting bandwidth and power of the WCD via the design of an acoustic echo cancellation (AEC)-aided CS-based SSP technique. From the experimental results, the voice data can be compressed and well-recovered efficiently via the proposed CS-based SSP system.

Keywords: Compressive Sensing；Speech Signal Processing；Wearable Computing；IPTV；Voice Control；AEC

Kuei-Hong, Lin (✉),Cheng-Hsun Lin, Kuo-Huang Chung and Kai-Shun Lin,
Information and Communications Research Laboratories,
Industrial Technology Research Institute of Taiwan, R.O.C.,
Hsinchu, Taiwan,
e-mail: kueihong990316@itri.org.tw

1 Introduction

The IPTV [1,2] services and applications have been integrated successfully which provide the unlimited enjoyment from the plentiful contents to the users. In these years, the IPTV technique has been integrated to the set-top box, smart TV, and the portable devices widely. For providing a user-friendly entertainment environment, the suitable user interface (UI) and UX of the IPTV has been studied in some researches [3,4]. These researches try to use the endpoint products to control the IPTV such as smart phone, tablet, keyboard, mouse, IP-camera, etc. However, these approaches are not comfortable and convenient since the users should hold something (e.g., smart phone, tablet, keyboard, and mouse) or be monitored (e.g., IP-camera) continuously. Finally, the users still have to select the traditional remote controller to interact with their TV, but the UX is not good. Using the controller of last century to control the modern TV is an interesting problem, and it is lamentable that the modern technologies cannot help you anything, especially in TV control.

In 2013, an innovative WCD named Google Glass (GG) is published. The device can be controlled using "voice commands", and do image recognition. With GG, the users' behaviors are integrated with GG except they are sleeping. The users can shuttle between the real and virtual world which provides the completely new UI/UX. In our research, we connect the IPTV with WCDs (ex: GG or iWatch), and try to imagine the scenario that the IPTV's services can be controlled by these WCDs one day. Based on this idea, we attempt to find the problems that may be happened in the future. By referring to the design of GG (http://www.google.com/glass/start/), we discovered that the GG keeps to record the background voice, and waits to "listen" the keywords "Ok, glass!". After recognizing the keywords, the miraculous powers of GG will be triggered. However, if the scenario is happened in the indoor entertainment environment of IPTV, it may not work since the background voice of the TV shows is too load, and the user's voice cannot be obtained by GG. Besides, the recorded background voice needs to be transmitted to the cloud continuously which wastes the bandwidth and power of GG.

For saving the bandwidth and power, the CS technique is applied [5] in this study which is a signal processing technique for acquiring and reconstructing the signal by finding the solutions of the underdetermined linear systems. The CS technique applies the sparsity of the signal in some domain, and let the signal to be determined via relatively few measurements. The CS technique is a great originality of signal processing which has been applied to many fields such as magnetic resonance imaging (MRI) [6], image processing [7], and the innovative analog to digital converter (ADC) named as analog to information converter (AIC) [8], etc.

This study is organized as follows. In section II, the brief introduction of the CS technique is given. In section III, the signal and signal model of the recorded voice in indoor IPTV environment will be formulated. In section IV, the proposed CS-based SSP system is discussed. In section V, the experimental environment is built for evaluating the proposed CS-based SSP system. In section VI, the experiment results will be given and discussed. In section VII, the concluding comments are made.

2 Brief Introduction of the Compressive Sensing

The CS technique is comprised of three steps, i.e., sparse representation of the signal, random measurement of the signal, and signal reconstruction. For example, let the real signal $\mathbf{x} \in R^N$ be sparse in the orthogonal basis $\mathbf{\Psi} = \{\Psi_0, \Psi_1, \cdots, \Psi_{N-1}\}$ represented as

$$\mathbf{x} = \mathbf{\Psi\theta} \quad (1),$$

where $\mathbf{x} = [x(0)\ x(1) \cdots x(N-1)]^T$, $\mathbf{\Psi} = [\mathbf{\Psi}_0\ \mathbf{\Psi}_1 \cdots \mathbf{\Psi}_{N-1}]$ is the matrix which contains the vector of the orthogonal basis in each column, and $\mathbf{\theta} = [\theta(0)\ \theta(1) \cdots \theta(N-1)]^T$ is the vector of the coefficients for the signal $\mathbf{x}$ in $\mathbf{\Psi}$. Note that $\mathbf{\theta}$ is a K-sparse vector, that is, the number of non-zero elements K is smaller than N, and K should satisfy $K \leqq N/2$. After making the sparse representation of $\mathbf{x}$ in (1), the random measurement of $\mathbf{x}$ is expressed as

$$\begin{aligned} \mathbf{y} &= \mathbf{\Phi x} \\ &= \mathbf{\Phi\Psi\theta} \end{aligned} \quad (2),$$

where $\mathbf{\Phi}$ is a $M \times N$ measurement matrix (MM), and $\mathbf{y} = [y(0)\ y(1) \cdots y(M-1)]^T$ is the measured. The measured signal can be seem as the compressed signal since M is smaller than N. Note that the MM in (2) should fit the restricted isometry property (RIP) [5], and the compressed signal (2) can be reconstructed via the basis pursuit (BP)-based algorithm (L1- minimization) [5] or the orthogonal matching pursuit (OMP) [9]. The reconstructed signal is defined as

$$\hat{\mathbf{x}} = \mathbf{\Psi}\hat{\mathbf{\theta}} \quad (3),$$

where $\hat{\mathbf{x}}$ is the reconstructed signal, and $\hat{\mathbf{\theta}}$ is the estimated T-sparse coefficients' vector which is acquired by the reconstructing algorithms.

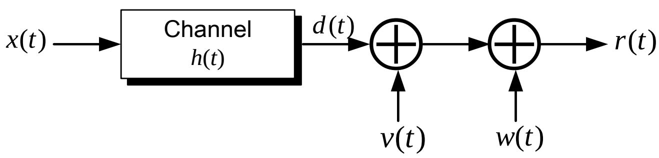


Fig. 1 Signal and system model of the recorded voice

3 Problem Formulation and Description

In general, the WCD may keep on recording the background sound, and it will await to be voice controlled anytime. As shown in Fig. 1, the signal and system model of the recorded voice $r(t)$ can be formed as

$$\begin{aligned} r(t) &= s(t) \otimes h(t) + v(t) + w(t) \\ &= d(t) + v(t) + w(t) \end{aligned} \quad (4),$$

where $s(t)$ is the audio source of the TV's content, the operator $\otimes$ denotes the convolution, $h(t)$ is the channel impulse response (CIR), $v(t)$ is the user's voice, $w(t)$ is the noise term, and $d(t)$ is the convolution of $x(t)$ and $h(t)$ represented as

$$d(t) = \sum_{l=0}^{L-1} h_l s(t - \tau_l) \quad (5),$$

where L is the channel's order, $\{h_l\}_{l=0}^{L-1}$ is the path attenuation, and $\{\tau_l\}_{l=0}^{L-1}$ is path excess delay. After sampling $r(t)$ by the analog to digital convertor (ADC), the sampled $r(t)$ would be transmitted to the cloud side for further analyzing which is expressed as

$$\begin{aligned} r(n) &= \mathbf{hs}(n) + v(n) + w(n) \\ &= d(n) + v(n) + w(n) \end{aligned} \quad (6),$$

where n is the sampling index, $\mathbf{h} = [h(0)\ h(1) \cdots h(L-1)]$ is the sampled CIR, $\mathbf{s}(n) = [s(n)\ s(n\text{-}1) \cdots s(n\text{-}L+1)]^T$ is the audio source's vector of the TV's content which is composed of the sampled $s(t)$, $d(n)$ is the echo-combined term. The vector-matrix representation of $r(n)$ is

$$\begin{aligned} \mathbf{r}(n) &= \mathbf{S}(n)\mathbf{h}^T + \mathbf{v}(n) + \mathbf{w}(n) \\ &= \mathbf{d}(n) + \mathbf{v}(n) + \mathbf{w}(n) \end{aligned} \quad (7),$$

where $\mathbf{r}(n)=[r(n)\ r(n+1)\cdots r(n+N-1)]^T$ is the vector of the recorded voice, $\mathbf{S}(n)=[\mathbf{s}(n)^T\ \mathbf{s}(n+1)^T\cdots\mathbf{s}(n+N-1)^T]^T$ is a $N \times N$ convolution matrix of the audio source, $\mathbf{v}(n)=[v(n)\ v(n+1)\cdots v(n+N-1)]^T$ is the user's voice, $\mathbf{w}(n)=[w(n)\ w(n+1)\cdots w(n+N-1)]^T$ is the noise vector, and $\mathbf{d}(n)=[d(0)\ d(1)\cdots d(N-1)]^T$ is the vector of the echo-combined term. For transmitting $\mathbf{r}(n)$ in (7) continuously to be voice recognition on the cloud, it makes the waste of the bandwidth and power of the WCD. For example, if we are lying on the sofa and enjoying the IPTV shows lazily, and the Google Glass is worn for controlling the smart TV or set-top box such as searching metadata of the video contents or providing the recommendation of the playlist. In this scenario, the voltage (life time) of the Google Glass's battery may decrease quickly. Therefore, an efficient approach to transmit $r(n)$ in (6) is important for the WCD.

4 Proposed System

For saving the bandwidth and power of the WCD to increase the transmitting efficiency, the CS technique is applied to the proposed SSP system (Fig. 2) since it can compress $\mathbf{r}(n)$ in (7) effectively. With fewer data, the compressed $\mathbf{r}(n)$ can be transmitted via lower bandwidth and shorter time, that is, the power of the WCD can be saved simultaneously. Hence, the battery life of the WCD is extended.

In Fig. 2, the overall architecture of the proposed CS-based SSP system is taken into the client and cloud site, and the proposed system will be introduced later.

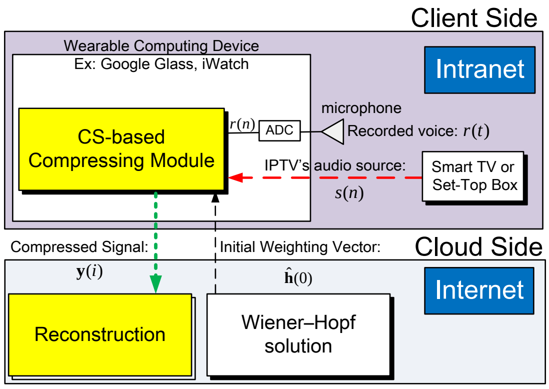


Fig. 2 Proposed CS-based SSP system

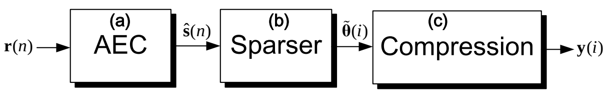


Fig. 3 Proposed CS-based SSP module

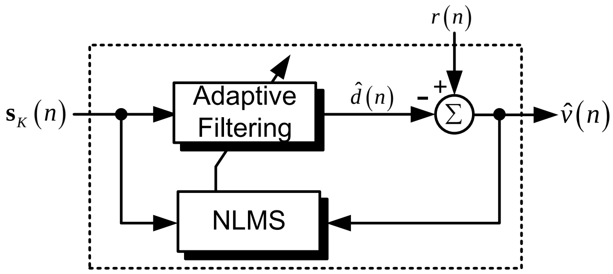


Fig. 4 Acoustic Echo Cancellation (AEC)

4.1 Client Side

In client side, it contains a smart TV or Set-Top box for providing IPTV services, and a WCD with the proposed AEC-aided CS-based SSP module inside.

In the smart TV or Set-Top box, an application would be installed to acquire the audio source of the content. For being applied to WCD's CS-based compressing module, the acquired audio source would be down sampled to achieve the same sampling rate as the WCD, and the down sampled audio source s(n) will be sent to the WCD's CS-based compressing module via the intranet simultaneously.

In the WCD, the background sound $r(t)$ in (4) would be recorded by the WCD's microphone continuously. After sampling $r(t)$ by ADC, the sampled background sound r(n) in (6) would be inputted to the CS-based compressing module for further processing with s(n).

In Fig. 2, the CS technique is applied to compress $\mathbf{r}(n)$ in (6) and (7) for increasing the transmitting efficiency. However, if the MM Φ is applied to compress r(n) directly, the convolution matrix S(n) will destroy the orthogonality of the MM Φ since each row and column are correlated. Hence, the echo-combined term or vector must be eliminated from (6) and (7), that is, the effect caused by CIR should be taken away.

In Fig. 3, the proposed CS-based compressing module has three steps inside. In step (a), the acoustic echo cancellation (AEC) method as shown in Fig. 4 is applied to eliminate the echo-combined term in (6) and (7), and achieve the estimate of the user's voice expressed as

$$\hat{v}(n) = r(n) - \hat{d}(n) \qquad (8),$$

where n is the sample index, $\hat{d}(n) = \hat{\mathbf{h}}(n)\mathbf{s}_D^T(n)$ is the estimate of the echo-combined term, and $\hat{\mathbf{h}}(n) = \left[\hat{h}(0)\ \hat{h}(1)\cdots\hat{h}(D-1)\right]$ is the estimate of the CIR acquired by normalized least mean square (NLMS) algorithm [10] expressed as

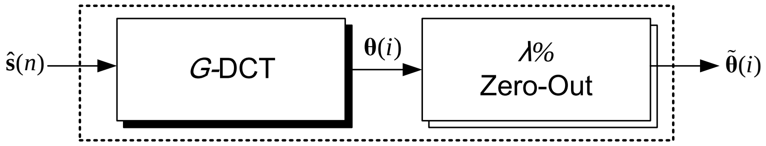


Fig. 5 Sparser

$$\hat{\mathbf{h}}(n) = \mathbf{h}(n-1) + \frac{\mu \hat{v}(n-1)\mathbf{s}_D(n-1)}{\left\|\mathbf{s}_D(n-1)\right\|^2} \quad (9),$$

where D is tap's length of the adaptive filter in AEC which is assumed to be longer than the channel's order L, μ is step's size, $\mathbf{s}_D(n-1) = \left[s(n-1)\ s(n-2) \cdots s(n-D)\right]$ is the vector of the inputted IPTV's audio source. Note that the IPTV's audio source (44.1 kHz) is pre-down sampled to the same sampling rate as the WCD. For providing better user experience, the filter's weights is hoped to be converged rapidly. Therefore, the cloud side is applied to compute the Wiener–Hopf solution [11] to be the initial weighting vector $\hat{\mathbf{h}}(0)$ of the adaptive filer since the matrix computation is too complex to be operated in the WCD. The Wiener–Hopf solution is given as

$$\hat{\mathbf{h}}(0) = \left[\mathbf{R}(n)^{-1}\mathbf{p}(n)\right]^T \quad (10),$$

where $\mathbf{R} = E\left[\mathbf{s}_D(n)^T \mathbf{s}_D(n)\right]$ is the auto-correlation matrix, and $\mathbf{p}(n) = \left[\mathbf{s}_K(n) r(n)\right]^T$ is the cross-correlation vector.

After removing the CIR's effect, the estimate of $v(n)$ in (8) will be compressed via the CS technique. For applying the CS technique to compress $\hat{v}(n)$, it should be transformed to the sparse domain in the step (b) of the proposed CS-based SSP module. In Fig. 5, the G-points discrete cosine transform (DCT) [12] is applied to make the speech signal be sparser and acquire relatively few coefficients. According to (1), the G-DCT process can be represented as

$$\boldsymbol{\theta}(i) = \boldsymbol{\psi}_G \hat{\mathbf{v}}(i) \quad (11),$$

where $\boldsymbol{\theta}(i) = \left[\theta_i(0)\ \theta_i(1) \cdots \theta_i(G-1)\right]^T$ is the i^{th} G-DCT coefficients' vector, $\boldsymbol{\psi}_G$ is defined as a GxG DCT matrix, and $\hat{\mathbf{v}}(i) = \left[v(G\times i+0)\ v(G\times i+1) \cdots v(G\times(i+1)-1)\right]^T$ is the i^{th} inputted vector of the user's voice's estimate. Note that the most of the energy is carried out by the largest DCT coefficients [12] as shown in Fig. 6(b). For making the DCT coefficients be sparser, the DCT coefficients are sorted, and the λ% smaller DCT coefficients are set to zeros, that is, the number non-zeros elements is

$$K = \left(1 - \lambda \times 10^{-2}\right) G \le G/2 \qquad (12),$$

where K can be seem as the order of (11), and the zero-out ratio λ is suggested setting as 83 in [14]. The i^{th} λ% zeroed-out DCT coefficients' vector is defined as

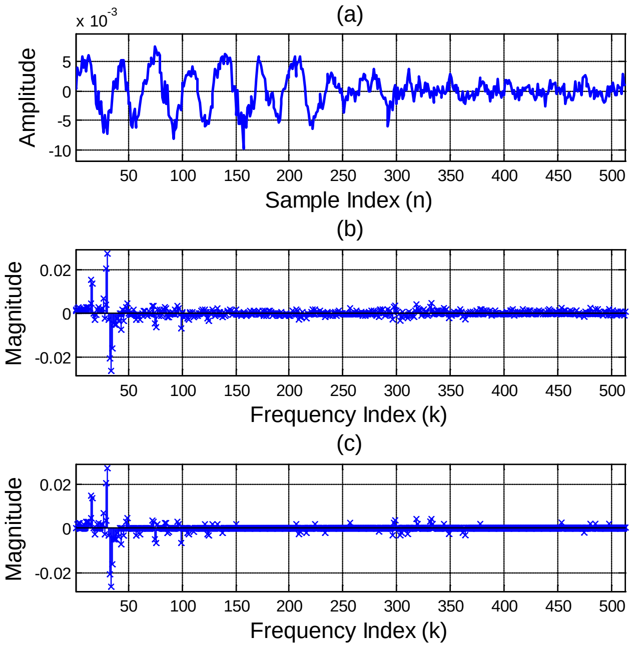


Fig. 6 Sparse procedure (a) $\hat{\mathbf{v}}(i)$: estimate of the user's voice: (b) $\boldsymbol{\theta}(i)$: DCT coefficients' vector (c) $\boldsymbol{\theta}_{\lambda}(i)$: 83% zeroed-out DCT coefficients

$\boldsymbol{\theta}_{\lambda}(i)$. As shown in Fig. 6(c), the zero-out ratio λ is set as 83 for eliminating the smaller coefficients of Fig. 6(b), and the signal to noise ratio (SNR) is set as 30dB in this example.

In the step (c) of the proposed CS-based SSP module, the λ% zeroed-out DCT coefficients $\boldsymbol{\theta}_{\lambda}(i)$ will be measured and compressed by the MM $\boldsymbol{\Phi}$ representation as

$$\mathbf{y}(i) = \boldsymbol{\Phi}\boldsymbol{\theta}_{\lambda}(i) \qquad (13),$$

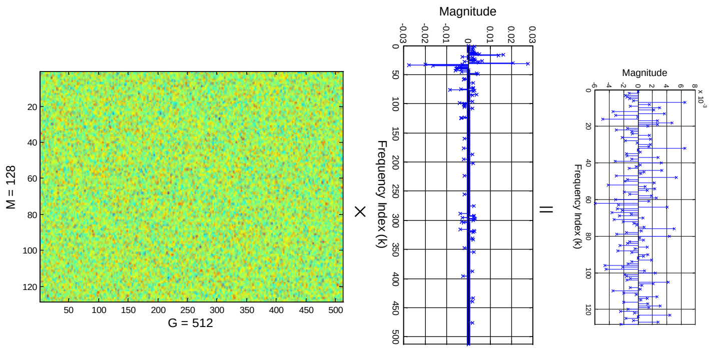
where $\mathbf{y}(i) = [y_i(0)\ y_i(1) \cdots y_i(M-1)]^T$ is the i^{th} measured and compressed G-DCT coefficients' vector, and $\boldsymbol{\Phi}$ is a $M \times G$ matrix. Note that the MM $\boldsymbol{\Phi}$ is assumed to be fitted the RIP [5], and the magnitude of M [5] should satisfy

$$M \geq CK \log\left(G/K\right) \quad (14),$$

where C<1 is a constant which is suggested setting as 0.28 in [13], and the K is the order of $\boldsymbol{\theta}_\lambda(i)$ defined in (12). The magnitude of M can be computed by replacing K in (12), and the recomputed M is given as

$$M \geq C\left(1-\lambda\times10^{-2}\right)G\log\left(\frac{1}{1-\lambda\times10^{-2}}\right) \quad (15),$$

As shown in Fig. 7, the $\boldsymbol{\theta}_\lambda(i)$ of Fig 6(c) is compressed by the MM with the magnitude of M=128 which satisfies $M \geqq 43$. The compressing ratio is defined as $\beta = \lceil G/M \rceil$. Thus, β is 4 in this example, that is, the transmitting power and bandwidth can be saved efficiently.



$$\boldsymbol{\Phi} \quad \times \quad \boldsymbol{\theta}_\lambda(i) \quad = \quad \mathbf{y}(i)$$

Fig. 7 Example of the Compressing Procedure

4.2 Cloud Side

After the compressing procedure in (13), the compressed signal would be transmit to cloud side for further reconstruction and recognition by parallel computation. For reconstructing the compressed signal, the L1-minimization algorithm [5] is employed to solve the nondeterministic and convex optimization problem given as

$$\hat{\boldsymbol{\theta}}_\lambda(i) = \min_{\hat{\boldsymbol{\theta}}_\lambda(i)} \left\| \hat{\boldsymbol{\theta}}_\lambda(i) \right\|_1 \text{ subject to } \mathbf{y}(i) = \boldsymbol{\Phi}\hat{\boldsymbol{\theta}}_\lambda(i) \quad (16),$$

and the recovered signal can be acquired by making the inverse DCT to the estimated coefficients' vector in (16) expressed as

$$\tilde{\mathbf{v}}(i) = \boldsymbol{\psi}_G^{-1}\hat{\lambda}(i) \qquad (17),$$

where $\tilde{\mathbf{v}}(i)$ is the reconstruction of $\hat{\mathbf{v}}(i)$, $\boldsymbol{\psi}_G^{-1}$ is the inverse DCT matrix. Note that each $\tilde{\mathbf{v}}(i)$ can be reconstructed in difference machines via the parallel computing, and the recombination can be operated via the index i of each $\hat{\mathbf{v}}(i)$. As shown in Fig. 8, the user's sound $\hat{\mathbf{v}}(i)$ of Fig.6 (a) is recovered by L1-minimization algorithm [5] which verified that the proposed CS-based SSP system can save the transmitting power and bandwidth efficiently, and gets the well-recovered signal.

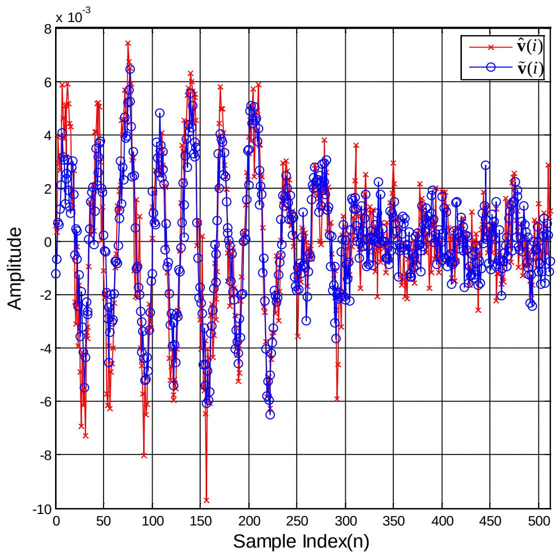


Fig. 8 Reconstruction of $\hat{\mathbf{v}}(i)$

5 Experiment Setup

For testing the proposed CS-based SSP system under an indoor environment, we modified the MATLAB's source code from "http://www.mathworks.com/help/dsp/examples/acoustic-echo-cancellation-aec.html" which is an experiment about the AEC technique. In this experiment, the user is making the audio teleconferencing whose speech is affected by the indoor (room) CIR as shown, and the AEC is applied to operate the echo cancellation.

As shown in **Table 1**, the recorded voice $r(t)$ is sampled under the 8k Hz, where the sampled $r(t)$ can be seem as the combination of the echo-effected TV's audio source $d(n)$ with the user's voice $v(n)$. Note that the indoor CIR for convoluting with TV's audio source $s(n)$ is generated randomly in each simulation, and the sampled CIR's length (L) is 1024.

Table 1 Simulation Parameters

Parameters	Settings
Simulation times	200
Speech sampling rate	8kHz
Signal to noise ratio (SNR)	[0 5 10 15 20 30] dB
Channel's order (L)	1024
Tap's length (D) of AEC's NLMS filter	1024
NLMS adaptive filter's step's size (μ)	0.05
Zero-out ratio (λ)	83
G-DCT	512
Magnitude of M of $\mathbf{\Phi}$	[32 43 48 64 128 256]
Type of $\mathbf{\Phi}$	Random MM
Compressing ratio (β)	[16 12 11 8 4 2]

In the proposed CS-based SSP module of the WCD (Fig. 3), the tap's length (D) of AEC's NLMS filter is 1024, and the filter's step's size (μ) is 0.05. In the sparser (Fig. 5), the G-DCT with G=512 is operated, and the zero-out ratio λ = 83. In the compressing procedure (step (c) of Fig. 3), we set different magnitude of M of MM $\mathbf{\Phi}$, where the setting of M should be satisfied

$$M \geq \left\lceil \begin{array}{l} 0.28\times\left(1-83\times10^{-2}\right)\times512 \\ \times\log\left(\dfrac{1}{\left(1-83\times10^{-2}\right)}\right) \end{array} \right\rceil = 43 \quad (18),$$

and the tested M are [32 43 48 64 128 256] with corresponded compressing ratio β=[64 27 16 8 4 2], respectively. Besides, the Pearson's correlation is applied to evaluate the performance of the proposed CS-based SSP system. The similarity or precision of the recovered user's voice $\tilde{v}(n)$ can be known by computing the Pearson's correlation coefficient defined as

$$\rho\left(v,z\right)=\frac{\sum_{n=0}^{N'-1}\left(v(n)-\overline{v}\right)\left(z(n)-\overline{z}\right)}{\sqrt{\sum_{n=0}^{N'-1}\left(v(n)-\overline{v}\right)^2}\sqrt{\sum_{n=0}^{N'-1}\left(z(n)-\overline{z}\right)^2}} \quad (19),$$

where $-1 \leqq \rho(v,z) \leqq 1$, N' is the total number of the tested samples, $v(n)$ is the user's voice, and $z(n)$ is the signal for comparing with $v(n)$. In our experiment, the $z(n)$ can be $\hat{v}(n)$ in (8) and $\tilde{v}(n)$, where $\tilde{v}(n)$ is the recovered $\hat{v}(n)$. Note that

$$\bar{v} = \frac{1}{N'} \sum_{n=0}^{N'-1} v(n) \tag{20}$$

and

$$\bar{z} = \frac{1}{N'} \sum_{n=0}^{N'-1} z(n) \tag{21}$$

are the averages of $v(n)$ and $z(n)$, respectively.

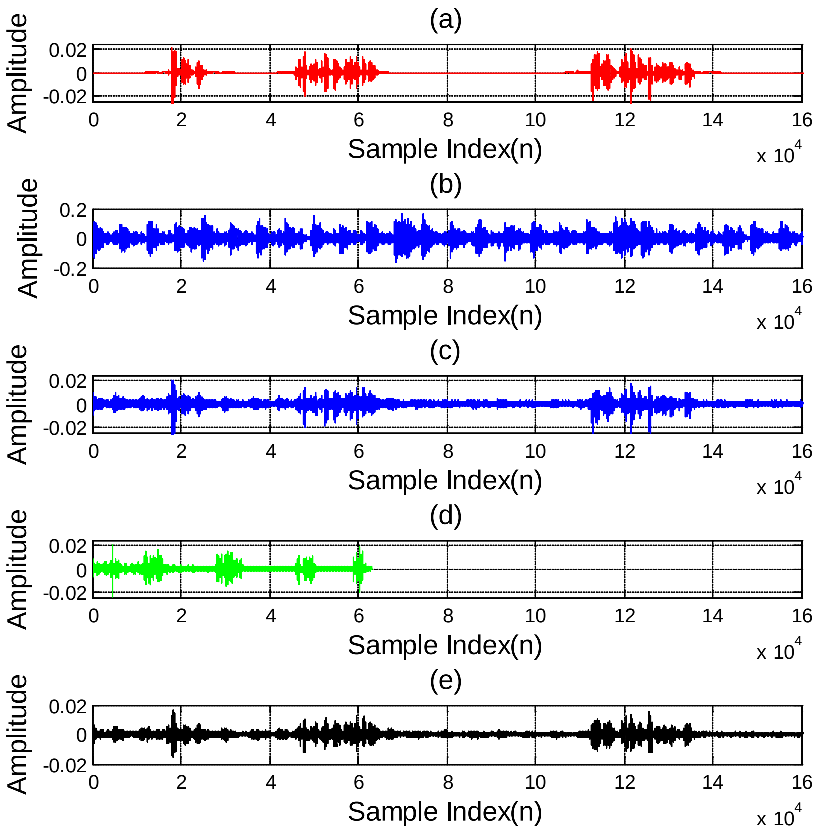


Fig. 9 Overall test of the proposed SSP system (a) $v(n)$: user's voice (b) $r(n)$: recorded voice (c) $\hat{v}(n)$: user's voice estimated by AEC (d) $\mathbf{y}(i)$: compressed the zeroed-out DCT coefficients $\boldsymbol{\theta}_{\lambda}(i)$ by MM $\boldsymbol{\Phi}$ (e) $\tilde{v}(n)$: user's voice reconstructed by L1-minimization [5]

6 Experiment Results

The overall test of the proposed CS-based SSP system is given in Fig.10 which is under SNR = 30dB, and the magnitude of M = 128 In Fig. 10(b), the user's voice v(n) (Fig. 9(a)) is mixed with the echo-combined term $d(n)$ then $d(n)$ is eliminated by AEC as shown in Fig. 9(c) for acquiring the estimate of the user's voice $\hat{v}(n)$ in (8). In Fig. 9(d), the zeroed-out DCT coefficients $\boldsymbol{\theta}_{\lambda}(i)$ is compressed by MM $\boldsymbol{\Phi}$ in (13). In Fig. 9(e), the recovered user's voice $\tilde{v}(n)$ is acquired by (16) and (17). By comparing the recovered user's voice $\tilde{v}(n)$ in Fig. 9(e) with the user's voice $v(n)$ in Fig. 9(a), and the result of $\tilde{v}(n)$ is similar to $v(n)$ as shown in Fig. 10.

In Fig.11, the proposed CS-based SSP system is tested under different compressing ratios (β). By contrast with the **Table 2** [16], the proposed SSP system can achieve moderately correlation (ρ=0.4 to 0.59) while the compressing ratio (β) is higher than 12, and it can achieve strong correlation (ρ=0.6 to 0.79) or very strong correlation (ρ=0.8 to 1.0) while the compressing ratio (β) is higher than 4. Besides, the results proves that the magnitude of M should be set higher than 43 computed in (18), that is, β should higher than 12 for acquiring better results, or the correlation between the user's voice $v(n)$ and the recovered voice $\tilde{v}(n)$ would be lower as shown in the case: M=32 and β=16.

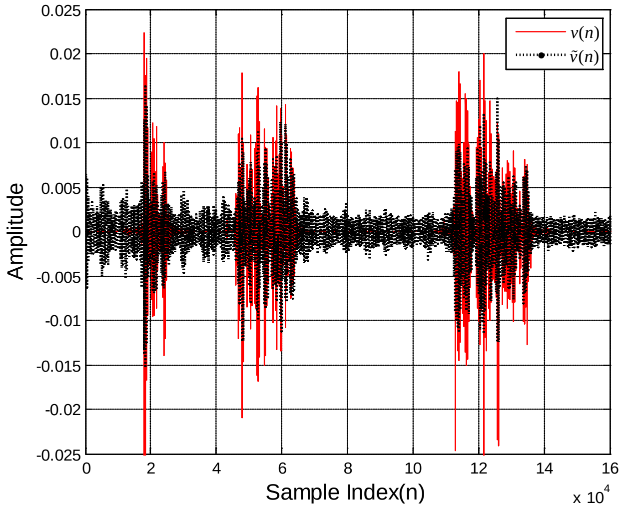


Fig. 10 Compare $v(n)$ to $\tilde{v}(n)$

Table 2 Definition of Correlation's Strength

Correlation's Strength	Absolute Value of ρ
Very Strong	0.80~1.0
Strong	0.60~0.79
Moderate	0.40~0.59

Weak□	0.20~0.39
Very Weak	0.00 ~0.19

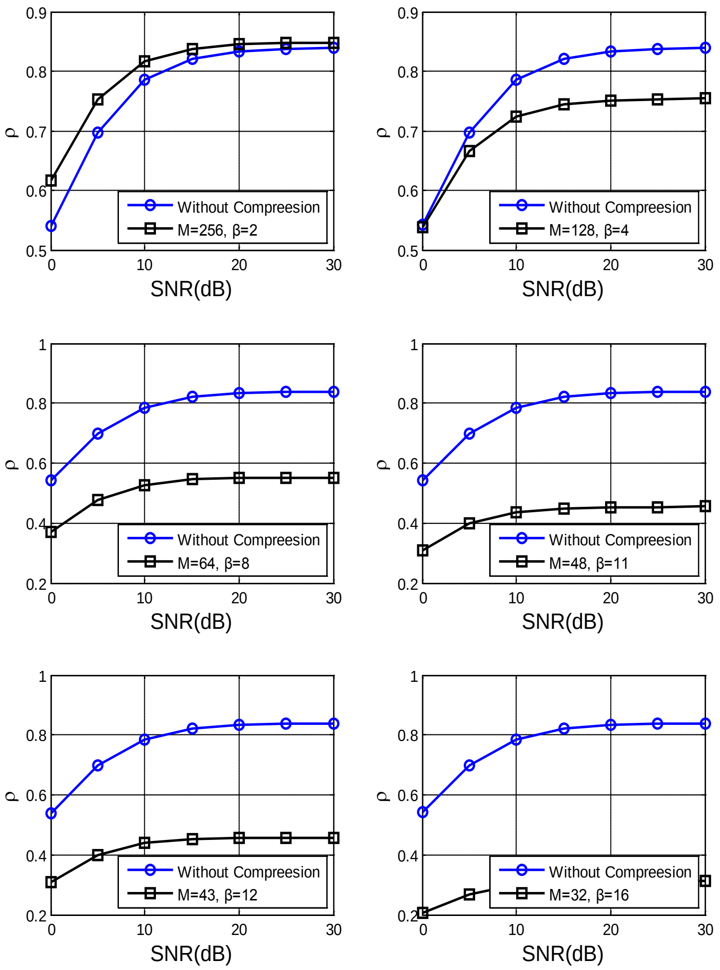


Fig. 11 Performance evaluation under different compressing ratios (β)

7 Conclusion

A CS-based SSP system is proposed for saving the WCD's power and bandwidth in this study. With the AEC-aided design, the compressing procedure would not be affected by the CIR, that is, the orthogonally of MM would not be destroyed. From the experimental results, the proposed CS-based SSP system can compress and recover the speech data efficiently, and the recovered speech is highly correlated with the user's voice. Thus, the power and bandwidth of the WCD can be saved. For reducing the computational complexity, the refinements of the proposed CS-based SSP system would be studied continuously.

Acknowledgment

This research is supported by the Information and Communications Research Laboratories (ICL), Industrial Technology Research Institute (ITRI), Taiwan, Republic of China (ITRI Grant Project Code C352BR2100)

References

1. Vasudevan, S.V., Xiaomei, L. Kollmansberger, K. (2008) *IPTV Systems, Standards and Architectures: Part II - IPTV Architectures for Cable Systems: An Evolutionary Approach*, IEEE Communications Magazine, Vol. 46, No. 5, 102~109.
2. Maisonneuve, J., et al. (2009) *An Overview of IPTV Standards Development*, IEEE Transactions on Broadcasting, v.55, n.2, 315-328.
3. Jana, R., Chen, Y.-F., Gibbon, D., Huang, Y., Jora, S., Murray, J., Wei, B. (2007) *Clicker - an IPTV remote control in your cell phone*, IEEE International Conference on Multimedia and Expo, 1055-1058.
4. Lin, C., Chen, M. (2005) *On controlling digital TV set-top-box by mobile devices via IP network*, Proceedings of the 7th IEEE International Symposium on Multimedia, 1-8.
5. Candès, E. J., Wakin, M. B. (2008) *An Introduction To Compressive Sampling*, IEEE Signal Proc Mag., 25, 21-30.
6. Lustig, M., Donoho, D., Santos, J. and Pauly, J. (2008). *Compressed sensing MRI*, IEEE Signal Processing Magazine 27, 72–82.
7. Duarte, M. F., Davenport, M. A., Takhar, D., Laska, J. N., Sun, T., Kelly, K. F., Baraniuk, R. G. (2008) *Single-pixel imaging via compressive sensing*, IEEE Signal Process. Mag., vol. 25, 83 -91.
8. Kirolos, S. (2006) *Analog-to-information conversion via random demodulation*, Proc. IEEE Workshop Design, Applicat., Integrat. Software, 71 -74.
9. Tropp, J., Gilbert, A.C. (2007) *Signal recovery from partial information via orthogonal matching pursuit*, IEEE Trans. Inform. Theory, vol. 53, no. 12, 4655-4666.
10. Haykin, S.(2002) *Adaptive Filter Theory, 4th ed.*, Prentice Hall, 320–326.
11. Kay, S. M. (1993) *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 365-443.
12. Khayam, S.A. (2003) *The Discrete Cosine Transform(DCT):Theory and Application*, Department of Electrical & Computer Engineering Michigan State University.
13. Elda, Y.C., Kutyniok, G. (2012) *Compressed Sensing: Theory and Applications, 1st ed.*, Cambridge University Press, 1-64.
14. Ramadan, M.H.(2010) *Compressive Sampling of Speech Signals*, Department of Science in Electrical Engineering University of Pittsburgh, 35-36.
15. Breese, J.S., Heckerman, D., Kadie, C. (1998) *Empirical analysis of predictive algorithms for collaborative filtering*, Proceedings of the 14th conference on uncertainty in artificial intelligence, 43–52.

16. Evans, J.D. (1996) *Straightforward Statistics for the Behavioral Sciences*, Pacific Grove, CA: Brooks/Cole Publishing.