AVOIDING LOCAL TRAP IN NONLINEAR ACOUSTIC ECHO CANCELLATION WITH CLIPPING COMPENSATION

Hiroki Kuroda, Masao Yamagishi, Isao Yamada

Dept. of Communications and Computer Engineering, Tokyo Institute of Technology
Email:{kuroda,myamagi,isao}@sp.ce.titech.ac.jp

ABSTRACT

For the nonlinear acoustic echo cancellation, we present an adaptive learning of the saturation effect of the amplifier and the room propagation in terms of the hard-clipping and the FIR system. The conventional learning algorithms are based on a gradient descent method, i.e., rely on local information, which results in a major drawback that the estimation of the hard-clipping is trapped in local minima. In this paper, we solve this drawback by exploiting global information embodied as a set including the desired hard-clipping with high-probability. The proposed adaptive learning of the hard-clipping is designed to track the sets with a projection-based algorithm. In the adaptive learning of the FIR system, we propose the use of the Huber loss function for the robustness against the error in the estimation of the hard-clipping. Numerical examples show that the proposed algorithm is never trapped in the local minima and has an excellent steady-state behavior.

Index Terms— Nonlinear acoustic echo cancellation, memoryless nonlinearity, clipping compensation, adaptive filtering.

1. INTRODUCTION

Nonlinear acoustic echo cancellation (NLAEC) aiming to attenuate the nonlinear echo signal has become increasingly important because, e.g., today's telecommunication devices often include small amplifiers and loudspeakers introducing significant saturation nonlinearity [1–8]. The overall echo path is the cascade of the amplifier and the loudspeaker followed by the room propagation. The former ones and the latter one can be modeled by the hard-clipping and the finite impulse response (FIR) system, respectively [1–5]. A major goal of the NLAEC is to learn adaptively the threshold of the hard-clipping and the impulse response vector of the FIR system [1–5].

The conventional learning algorithms [1–5] are extension of the NLMS [9] and derived by a gradient descent method applied to the squared estimation residual. Dependency only on local information, i.e., the gradient, causes a major drawback, to which we refer as *local minima trapping*, that the threshold is never updated if its current estimate is larger than the amplitude of the input signal [2–4].

In this paper, we solve the local minima trapping by exploiting, as global information on the desired threshold, a *feasible threshold set* consisting of all thresholds satisfying that the estimation residual is less than a given constant. To use the set in the estimation of the hard-clipping, we provide its explicit representation as a union of closed intervals by using the piecewise linearity of the estimation residual. The proposed adaptive learning algorithm for the hard-clipping is derived by tracking the convex hull of the set based on the Projection Onto Convex Sets (POCS) [10–16]. Moreover, we present an efficient computation for these procedures, which has comparable computational complexity to the conventional algorithms. In the adaptive learning of the FIR system, we propose to use the Huber loss function [17] for the robustness against the error in the estimation of the hard-clipping in the initial stage.

Numerical examples show that the proposed adaptive learning of the hard-clipping is free from the local minima trapping, and the proposed simultaneous learning has the best steady-state behavior in the echo return loss enhancement.

The authors would like to thank Dr. Shimauchi for sending [4] and indicating the references [2,7]. This work was supported in part by JSPS Grants-in-Aid (24800022, B-21300091).

2. PRELIMINARIES

2.1. Nonlinear Acoustic Echo Cancellation

Let $\mathbb{N}$ and $\mathbb{R}$ respectively denote the sets of nonnegative integers and real numbers, and define $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$ and $\mathbb{R}_+ := \{x \in \mathbb{R} \mid x \geq 0\}$.

Consider a nonlinear echo path modeled by the cascade of the hard-clipping function followed by the FIR filter [1–5], namely, observation $(\boldsymbol{x}_k, d_k)$ at time $k \in \mathbb{N}^*$ is defined by

$$\boldsymbol{u}_k := \boldsymbol{\phi}_{\gamma^\star}(\boldsymbol{x}_k), \tag{1}$$

$$d_k := \boldsymbol{u}_k^t \boldsymbol{h}^\star + v_k, \tag{2}$$

where $\boldsymbol{x}_k := [x_k, \ldots, x_{k-N+1}]^t \in \mathbb{R}^N$ is the input (far-end speech) signal, $\boldsymbol{u}_k \in \mathbb{R}^N$ is the unknown clipped signal, $\boldsymbol{h}^\star \in \mathbb{R}^N$ is the room impulse response vector, d_k consists of the echo signal $y_k := \boldsymbol{u}_k^t \boldsymbol{h}^\star$ and noise $v_k \in \mathbb{R}$, $N \in \mathbb{N}^*$ is the length of the impulse response vector, and $(\cdot)^t$ denotes the transpose operation. The hard-clipping function with threshold value $\gamma \in \mathbb{R}_+$ is defined by[1]

$$\boldsymbol{\phi}_\gamma : \mathbb{R}^N \to [-\gamma, \gamma]^N, \boldsymbol{\phi}_\gamma(\boldsymbol{x}_k) := [\phi_\gamma(x_k), \ldots, \phi_\gamma(x_{k-N+1})]^t,$$

$$\phi_\gamma : \mathbb{R} \to [-\gamma, \gamma], \phi_\gamma(x) := \begin{cases} x, & \text{if } |x| \leq \gamma, \\ \gamma \operatorname{sgn}(x), & \text{otherwise.} \end{cases}$$

A major goal of the nonlinear acoustic echo cancellation (NLAEC) is a simultaneous approximation of $\gamma^\star$ and $\boldsymbol{h}^\star$ by their estimation sequences $(\gamma_k)_{k \in \mathbb{N}}$ and $(\boldsymbol{h}_k)_{k \in \mathbb{N}}$ with available data $(\boldsymbol{x}_i, d_i)_{i=1}^k$ and initial estimates $\boldsymbol{h}_0 \in \mathbb{R}^N$ and $\gamma_0 \in \mathbb{R}_+$. Note that different models are also found in the NLAEC (see, e.g., [6–8, 18, 19]).

2.2. Brief Review of Conventional Methods

The update of the conventional algorithms [1–5] for $\boldsymbol{h}_k$ and γ_k is an extension of the well-known NLMS [9] and derived from a wide-sense gradient descent method applied to the squared estimation residual function $e_k^2(\gamma, \boldsymbol{h})$:

$$\gamma_{k+1} = \gamma_k + (\mu_\gamma / w_\gamma) e_k(\gamma_k, \boldsymbol{h}_k) \boldsymbol{h}_k^t \nabla \boldsymbol{\phi}_{\gamma_k}(\boldsymbol{x}_k), \tag{3}$$

$$\boldsymbol{h}_{k+1} = \boldsymbol{h}_k + (\mu_{\boldsymbol{h}} / w_{\boldsymbol{h}}) e_k(\gamma_k, \boldsymbol{h}_k) \hat{\boldsymbol{u}}_k, \tag{4}$$

where μ_γ and $\mu_{\boldsymbol{h}}$ are the step-sizes with normalization constants w_γ and $w_{\boldsymbol{h}}$, $e_k : \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$ is defined by $e_k(\gamma, \boldsymbol{h}) := d_k - \boldsymbol{h}^t \boldsymbol{\phi}_\gamma(\boldsymbol{x}_k)$, $\hat{\boldsymbol{u}}_k := \boldsymbol{\phi}_{\gamma_k}(\boldsymbol{x}_k)$ is an estimated clipped signal, and $\nabla \boldsymbol{\phi}_{\gamma_k}(\boldsymbol{x}_k) \in$

[1] $\operatorname{sgn}(x) := x/|x|$ $(\forall x \in \mathbb{R} \setminus \{0\})$ and $\operatorname{sgn}(0) := 0$.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

$\mathbb{R}^N$ is the wide-sense gradient vector defined by $\nabla\boldsymbol{\phi}_{\gamma_k}(\boldsymbol{x}_k) := [\phi'_{\gamma_k}(x_k), \ldots, \phi'_{\gamma_k}(x_{k-N+1})]^t$ with the right derivative

$$\phi'_{\gamma_k}(x) := \frac{\partial}{\partial\gamma}\phi_\gamma(x)\bigg|_{\gamma=\gamma_k+0} = \begin{cases} 0, & \text{if } \gamma_k \geq |x|, \\ \operatorname{sgn}(x), & \text{otherwise.} \end{cases} \quad (5)$$

The main drawback of (3), to which we refer as the local minima trapping, is that γ_k is not updated regardless of $e_k(\gamma_k, \boldsymbol{h}_k)$ if γ_k is larger than the maximum amplitude of $\boldsymbol{x}_k$ because $\nabla\boldsymbol{\phi}_{\gamma_k}(\boldsymbol{x}_k)$ becomes $\mathbf{0}$ from (5). This drawback is partially discussed in [2–4] as an initialization issue.

3. MAIN CONTRIBUTIONS

3.1. Feasible Set for Hard-Clipping Estimation

To solve the local minima trapping, we introduce a feasible threshold set $S_k^{(\varepsilon_k)}$ consisting of all γ whose recent $r\ (\in \mathbb{N}^*)$ estimation residual $\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k)$ is suppressed sufficiently[2]:

$$S_k^{(\varepsilon_k)} := \left\{\gamma \in [0, \gamma_{\max}] \,\middle|\, \|\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k)\|_1 \leq \varepsilon_k \right\}, \quad (6)$$

where $\gamma_{\max}$ is an upper bound of threshold[3], $\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k) := [e_k(\gamma, \boldsymbol{h}_k), \ldots, e_{k-r+1}(\gamma, \boldsymbol{h}_k)]^t \in \mathbb{R}^r$, or

$$\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k) = \boldsymbol{d}_k - \boldsymbol{H}_k\bar{\boldsymbol{\phi}}_\gamma(\bar{\boldsymbol{x}}_k), \quad (7)$$

$\boldsymbol{d}_k := [d_k, \ldots, d_{k-r+1}]^t \in \mathbb{R}^r$, $\boldsymbol{H}_k \in \mathbb{R}^{r\times\bar{N}}$ is a Toeplitz matrix defined with $\boldsymbol{h}_k$ and represents the convolution, $\bar{\boldsymbol{\phi}}_\gamma(\bar{\boldsymbol{x}}_k) := [\phi_\gamma(x_k), \ldots, \phi_\gamma(x_{k-\bar{N}+1})]^t \in \mathbb{R}^{\bar{N}}$, $\bar{\boldsymbol{x}}_k := [x_k, \ldots, x_{k-\bar{N}+1}]^t \in \mathbb{R}^{\bar{N}}$, and $\bar{N} := N + r - 1$. A constant $\varepsilon_k \in \mathbb{R}_+$ determines the reliability of the set-membership: $\gamma^\star \in S_k^{(\varepsilon_k)}$. We choose the ℓ_1 norm in (6) for the robustness against the mismatch between $\boldsymbol{h}_k$ and $\boldsymbol{h}^\star$ and the impulsive noise occurred, e.g., in double-talk situations.

To use $S_k^{(\varepsilon_k)}$ in the estimation of $\gamma^\star$, we provide its explicit representation as a union of closed intervals $S_{k,i}^{(\varepsilon_k)}$. Our essential idea is based on observation in $r = 1$. That is, since the piecewise linearity of the hard-clipping $\boldsymbol{\phi}_\gamma(\boldsymbol{x}_k)$ w.r.t. γ implies that of $e_k(\gamma, \boldsymbol{h}_k)$, the set $S_k^{(\varepsilon_k)}$ becomes a union of closed intervals.

Proposition 1.

(a) (Piecewise linearity of $\bar{\boldsymbol{\phi}}_\gamma(\bar{\boldsymbol{x}}_k)$ and $\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k)$) Sort the absolute value of entries in $\bar{\boldsymbol{x}}_k$ into $(\gamma_{k,i})_{i=1}^{\bar{N}}$ in non-decreasing order, and define $\gamma_{k,0} := 0$ and $\gamma_{k,\bar{N}+1} := \gamma_{\max}$. Then, $\bar{\boldsymbol{\phi}}_\gamma(\bar{\boldsymbol{x}}_k)$ satisfies[4]

$$\bar{\boldsymbol{\phi}}_\gamma(\bar{\boldsymbol{x}}_k) = \gamma\boldsymbol{f}_i(\bar{\boldsymbol{x}}_k) + \boldsymbol{g}_i(\bar{\boldsymbol{x}}_k) \ \ (\gamma \in [\gamma_{k,i}, \gamma_{k,i+1}]), \quad (8)$$

where $\boldsymbol{f}_i(\bar{\boldsymbol{x}}_k) := \left(\bar{\boldsymbol{\phi}}_{\gamma_{k,i+1}}(\bar{\boldsymbol{x}}_k) - \bar{\boldsymbol{\phi}}_{\gamma_{k,i}}(\bar{\boldsymbol{x}}_k)\right)/(\gamma_{k,i+1} - \gamma_{k,i})$ and $\boldsymbol{g}_i(\bar{\boldsymbol{x}}_k) := \bar{\boldsymbol{\phi}}_{\gamma_{k,i}}(\bar{\boldsymbol{x}}_k) - \gamma_{k,i}\boldsymbol{f}_i(\bar{\boldsymbol{x}}_k)$. Eqs. (7) and (8) imply the piecewise linearity of $\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k)$, i.e.,

$$\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k) = \boldsymbol{b}_{k,i} - \gamma\boldsymbol{a}_{k,i} \ \ (\gamma \in [\gamma_{k,i}, \gamma_{k,i+1}]), \quad (9)$$

where $\boldsymbol{a}_{k,i} := \boldsymbol{H}_k\boldsymbol{f}_i(\bar{\boldsymbol{x}}_k)$ and $\boldsymbol{b}_{k,i} := \boldsymbol{d}_k - \boldsymbol{H}_k\boldsymbol{g}_i(\bar{\boldsymbol{x}}_k)$.

(b) (Explicit representation of $S_k^{(\varepsilon_k)}$) From (a), we can decompose $S_k^{(\varepsilon_k)}$ into the closed intervals $S_{k,i}^{(\varepsilon_k)}$, i.e.,

$$S_k^{(\varepsilon_k)} = \textstyle\bigcup_{i=0}^{\bar{N}} S_{k,i}^{(\varepsilon_k)},$$
$$S_{k,i}^{(\varepsilon_k)} = \left\{\gamma \in [\gamma_{k,i}, \gamma_{k,i+1}] \,\middle|\, \|\boldsymbol{b}_{k,i} - \gamma\boldsymbol{a}_{k,i}\|_1 \leq \varepsilon_k \right\}. \quad (10)$$

[2] The ℓ_1 norm of $\boldsymbol{x} \in \mathbb{R}^r$ is defined by $\|\boldsymbol{x}\|_1 := \sum_{i=1}^r |x_i|$.

[3] For the simplicity of notations, we here assume the existence of $\gamma_{\max}$. This assumption can be relaxed easily. Note that $\gamma_{\max}$ is set to sufficient large value in our experiments.

[4] In this paper, for the simplicity, we assume that entries of $(\gamma_{k,i})_{i=0}^{\bar{N}+1}$ are different. This assumption can be relaxed easily.

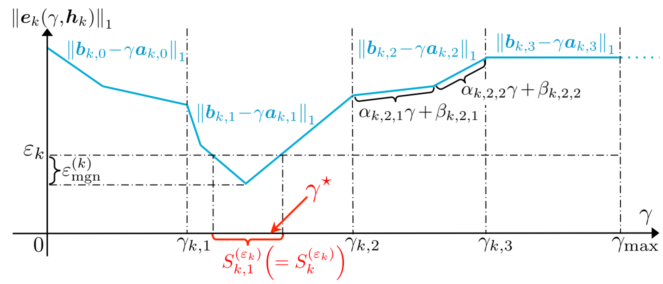


Fig. 1: An illustration of $S_{k,i}^{(\varepsilon_k)}$ in (10) and $S_k^{(\varepsilon_k)}$ and $\|\boldsymbol{e}_k(\boldsymbol{h}_k, \gamma)\|_1$ in (6) with $N = 2$ and $r = 2$.

Note that each closed interval $S_{k,i}^{(\varepsilon_k)}$ can be specified easily (see Remark 1(a)). An example of $S_{k,i}^{(\varepsilon_k)}$ is illustrated in Fig. 1. Combining Prop. 1 and the following Remark 1, we have an algorithm to obtain an explicit representation of $S_k^{(\varepsilon_k)}$, which is shown in Algorithm 1, with $\mathcal{O}\left(r\bar{N}\right)$ multiplications and $\mathcal{O}\left(\bar{N}\log_2\bar{N} + \bar{N}r\log_2 r\right)$ comparisons.

Remark 1.

(a) (Specification of the closed interval $S_{k,i}^{(\varepsilon_k)}$) We can specify $S_{k,i}^{(\varepsilon_k)}$ using analogous idea to Prop. 1. The piecewise linearity[5] of $\|\boldsymbol{b}_{k,i} - \gamma\boldsymbol{a}_{k,i}\|_1$ holds from the linearity of $\boldsymbol{b}_{k,i} - \gamma\boldsymbol{a}_{k,i}$, i.e.,

$$\|\boldsymbol{b}_{k,i} - \gamma\boldsymbol{a}_{k,i}\|_1 = \alpha_{k,i,j}\gamma + \beta_{k,i,j} \ \ (\gamma \in [p_{k,i,j}, p_{k,i,j+1}]), \quad (11)$$

where $(p_{k,i,j})_{j=0}^{r_{k,i}+1}$, $(\alpha_{k,i,j})_{j=0}^{r_{k,i}}$ and $(\beta_{k,i,j})_{j=0}^{r_{k,i}}$ can be calculated as shown in the step 3(i)(ii) of Algorithm 1. Therefore, $S_{k,i}^{(\varepsilon_k)}$ can be further decomposed into

$$S_{k,i}^{(\varepsilon_k)} = \textstyle\bigcup_{j=0}^{r_{k,i}} S_{k,i,j}^{(\varepsilon_k)}, \quad (12)$$
$$S_{k,i,j}^{(\varepsilon_k)} := \{\gamma \in [p_{k,i,j}, p_{k,i,j+1}] \mid \alpha_{k,i,j}\gamma + \beta_{k,i,j} \leq \varepsilon_k\}. \quad (13)$$

By computing the endpoints of the closed intervals $S_{k,i,j}^{(\varepsilon_k)}$ and combining them, $S_{k,i}^{(\varepsilon_k)}$ can be specified as[6]

$$S_{k,i}^{(\varepsilon_k)} = \left[\min_{j\in\{0,\ldots,r_{k,i}\}} \min S_{k,i,j}^{(\varepsilon_k)},\ \max_{j\in\{0,\ldots,r_{k,i}\}} \max S_{k,i,j}^{(\varepsilon_k)}\right]. \quad (14)$$

(b) (Choice of ε_k) To guarantee the nonemptiness of $S_k^{(\varepsilon_k)}$, we set $\varepsilon_k = \varepsilon_{\mathrm{mgn}}^{(k)} + \min_{\gamma\in[0,\gamma_{\max}]}\|\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k)\|_1$, where $\varepsilon_{\mathrm{mgn}}^{(k)} \in \mathbb{R}_+$ is a user-defined constant, and $\min_{\gamma\in[0,\gamma_{\max}]}\|\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k)\|_1$ is calculated as shown in the step 3(iii) of Algorithm 1.

(c) (Computation of $\boldsymbol{a}_{k,i}$ and $\boldsymbol{b}_{k,i}$ in $\mathcal{O}\left(r\bar{N}\right)$) By using the fact that $\boldsymbol{f}_i(\boldsymbol{x}_k) - \boldsymbol{f}_{i-1}(\boldsymbol{x}_k)$ has only one non-zero entry, $(\boldsymbol{a}_{k,i})_{i=0}^{\bar{N}}$ and $(\boldsymbol{b}_{k,i})_{i=0}^{\bar{N}}$ in (9) can be calculated as

$$\boldsymbol{a}_{k,i-1} = \boldsymbol{a}_{k,i} + \operatorname{sgn}\left(x_{k-\tau_k(i)+1}\right)\boldsymbol{c}_{\tau_k(i)}, \quad (15)$$
$$\boldsymbol{b}_{k,i} = \boldsymbol{b}_{k,i-1} - x_{k-\tau_k(i)+1}\boldsymbol{c}_{\tau_k(i)}, \quad (16)$$

for $i = 1, \ldots, \bar{N}$ with $[\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{\bar{N}}] := \boldsymbol{H}_k$, $\boldsymbol{a}_{k,\bar{N}} = \mathbf{0}$, $\boldsymbol{b}_{k,0} = \boldsymbol{d}_k$, and $\tau_k\colon \{1, \ldots, \bar{N}\} \to \{1, \ldots, \bar{N}\}$ satisfies $|x_{k-\tau_k(1)+1}| < \cdots < |x_{k-\tau_k(\bar{N})+1}|$ (see also footnote 4).

3.2. Adaptive Learning of Overall Echo Path

We derive an adaptive learning for the hard-clipping by tracking the

[5] The property has been shown in [20] using the relation to the weighted median search [21–23].

[6] For the empty set $\varnothing$, we let $\min\varnothing := \infty$ and $\max\varnothing := -\infty$.

Algorithm 1 Obtain an explicit representation of $S_k^{(\varepsilon_k)}$ in (6) by specifying $S_{k,i}^{(\varepsilon_k)}$ in (10) with $\bar{\boldsymbol{x}}_k$, $\boldsymbol{d}_k$, $\boldsymbol{H}_k$ in (7), $\gamma_{\max}$ in (6) and given $\varepsilon_{\mathrm{mgn}}^{(k)}$.

1: To clarify the ranges that $\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k)$ becomes linear as (9), sort the absolute value of entries in $\bar{\boldsymbol{x}}_k$ into $(\gamma_{k,i})_{i=1}^{\bar{N}}$ in non-decreasing order, and set $\gamma_{k,0} = 0$ and $\gamma_{k,\bar{N}+1} = \gamma_{\max}$.

2: To obtain a linear representation of $\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k)$ in each range, calculate $(\boldsymbol{a}_{k,i}, \boldsymbol{b}_{k,i})_{i=0}^{\bar{N}}$ in (9) (see its efficient computation (15), (16)).

3: Specify the closed interval $S_{k,i}^{(\varepsilon_k)}$ through the following steps.

(i) To clarify the ranges that $\left\|\boldsymbol{b}_{k,i} - \gamma \boldsymbol{a}_{k,i}\right\|_1$ becomes linear as (11), compute $(p_{k,i,j})_{j=0}^{r_{k,i}+1}$ by[7]

$$p_{k,i,j} := \begin{cases} \gamma_{k,i}, & \text{if } j = 0, \\ b_{k,i,\sigma_{k,i}(j)}/a_{k,i,\sigma_{k,i}(j)}, & \text{if } j = 1, \ldots, r_{k,i}, \\ \gamma_{k,i+1}, & \text{if } j = r_{k,i} + 1, \end{cases}$$

where $\sigma_{k,i} \colon \{1, \ldots, r_{k,i}\} \to \mathcal{J}_{k,i}$ is obtained by sorting $\{b_{k,i,j}/a_{k,i,j}\}_{j \in \mathcal{J}_{k,i}}$ in non-decreasing order so that $p_{k,i,1} \leq \cdots \leq p_{k,i,r_{k,i}}$, $r_{k,i} := |\mathcal{J}_{k,i}|$, and

$$\mathcal{J}_{k,i} := \{j \in \{1, \ldots, r\} | \exists \gamma \in (\gamma_{k,i}, \gamma_{k,i+1}) \text{ s.t. } b_{k,i,j} - \gamma a_{k,i,j} = 0\}.$$

(ii) To obtain a linear representation of $\left\|\boldsymbol{b}_{k,i} - \gamma \boldsymbol{a}_{k,i}\right\|_1$ in each range, calculate $(\alpha_{k,i,j}, \beta_{k,i,j})_{j=0}^{r_{k,i}}$ in (11) by[8]

$$\alpha_{k,i,j} := \begin{cases} -\boldsymbol{a}_{k,i}^t \overline{\mathrm{sgn}}_r(\boldsymbol{b}_{k,i} - \gamma_{k,i}\boldsymbol{a}_{k,i}, -\boldsymbol{a}_{k,i}), & \text{if } j = 0, \\ \alpha_j + 2|a_{k,i,\sigma_{k,i}(j)}|, & \text{otherwise}, \end{cases}$$

$$\beta_{k,i,j} := \begin{cases} \boldsymbol{b}_{k,i}^t \overline{\mathrm{sgn}}_r(\boldsymbol{b}_{k,i} - \gamma_{k,i}\boldsymbol{a}_{k,i}, -\boldsymbol{a}_{k,i}), & \text{if } j = 0, \\ \beta_{k,i,j} - 2|b_{k,i,\sigma_{k,i}(j)}|, & \text{otherwise}. \end{cases}$$

(iii) To guarantee the nonemptiness of $S_k^{(\varepsilon_k)}$, set $\varepsilon_k = \varepsilon_{\mathrm{mgn}}^{(k)} + \min_{\gamma \in [0, \gamma_{\max}]} \|\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k)\|_1$, or

$$\varepsilon_k = \varepsilon_{\mathrm{mgn}}^{(k)} + \min_{i \in \{0, \ldots, \bar{N}\}} \min_{j \in \{0, \ldots, r_{k,i}\}} \min_{\gamma \in [p_{k,i,j}, p_{k,i,j+1}]} \alpha_{k,i,j}\gamma + \beta_{k,i,j}.$$

(iv) Specify the closed interval $S_{k,i}^{(\varepsilon_k)}$ by (14) with calculating endpoints of $S_{k,i,j}^{(\varepsilon_k)}$ in (13).

Output $\bigcup_{i=0}^{\bar{N}} S_{k,i}^{(\varepsilon_k)}$.

convex hull[9] of $S_k^{(\varepsilon_k)}$ with a time-varying extension of the Projection Onto Convex Sets (POCS) [10–16]:[10]

$$\gamma_{k+1} := (1 - \mu_\gamma)\gamma_k + \mu_\gamma P_{\mathrm{conv} S_k^{(\epsilon_k)}}(\gamma_k), \tag{17}$$

where $\mu_\gamma \in (0, 2)$ is the step size, and the metric projection onto the convex hull of $S_k^{(\epsilon_k)}$ is given by

$$P_{\mathrm{conv} S_k^{(\epsilon_k)}}(\gamma_k) = \begin{cases} \min S_k^{(\varepsilon_k)}, & \text{if } \gamma_k < \min S_k^{(\varepsilon_k)}, \\ \max S_k^{(\varepsilon_k)}, & \text{if } \gamma_k > \max S_k^{(\varepsilon_k)}, \\ \gamma_k, & \text{otherwise}, \end{cases}$$

where $\min S_k^{(\varepsilon_k)}$ and $\max S_k^{(\varepsilon_k)}$ can be calculated as $\min_{i \in \{0, \ldots, \bar{N}\}} \min S_{k,i}^{(\varepsilon_k)}$ and $\max_{i \in \{0, \ldots, \bar{N}\}} \max S_{k,i}^{(\varepsilon_k)}$ with (14) and Algorithm 1.

Note that (17) can be explained in the frame of the adaptive projected subgradient method (APSM) [14–16] applied to a time varying cost function $d\left(\gamma, \mathrm{conv} S_k^{(\varepsilon_k)}\right)$, which ensures the properties of the APSM [15] including the monotone approximation $|\gamma_{k+1} - \gamma_*| < |\gamma_k - \gamma_*|$ for any $\gamma_k \notin \mathrm{conv} S_k^{(\varepsilon_k)}$ and $\gamma_* \in \mathrm{conv} S_k^{(\varepsilon_k)}$.

For the adaptive learning of the FIR system, we define the time-varying cost function $\Theta_k \colon \mathbb{R}^N \to \mathbb{R}$ by

$$\Theta_k(\boldsymbol{h}) := \rho_k(e_k(\gamma_k, \boldsymbol{h})),$$

where the Huber loss function $\rho_k \colon \mathbb{R} \to \mathbb{R}$ is defined as

$$\rho_k(x) := \begin{cases} x^2/2, & |x| \leq \delta_k, \\ \delta_k|x| - \delta_k^2/2, & \text{otherwise}, \end{cases}$$

with the cut-off value $\delta_k > 0$. We adopt the Huber loss function for the robustness against the error in γ_k and the impulsive noise caused in double-talk situations [25–27]. Applying the adaptive proximal forward-backward splitting (APFBS) [27–32] to Θ_k, we have the update for the FIR system:

$$\boldsymbol{h}_{k+1} := \boldsymbol{h}_k - \mu_{\boldsymbol{h}} \min\{1, \delta_k/|e_k(\gamma_k, \boldsymbol{h}_k)|\}(\boldsymbol{h}_k - P_{\Pi_k}(\boldsymbol{h}_k)), \tag{18}$$

where $\mu_{\boldsymbol{h}} \in (0, 2)$, and P_{Π_k} is the metric projection onto $\Pi_k := \arg\min_{\boldsymbol{h} \in \mathbb{R}^N} |e_k(\gamma_k, \boldsymbol{h})|$ calculated as

$$P_{\Pi_k}(\boldsymbol{h}_k) = \boldsymbol{h}_k + \frac{e_k(\gamma_k, \boldsymbol{h}_k)}{\|\hat{\boldsymbol{u}}_k\|_2^2} \hat{\boldsymbol{u}}_k.$$

Properties of the APFBS including the monotone approximation are shown in [28], and its excellent performance is confirmed, e.g., in [27–34].

4. SIMULATION RESULT

We conduct the simulations to show the efficacy of the proposed adaptive learning algorithm. In all the experiments, $\gamma^\star$ in (1) is set to 1, and $\boldsymbol{h}^\star$ in (2) is taken from [35] with $N = 1024$. The parameters of the algorithms are shown in Table 1.

First, to focus on the adaptive learning of the hard-clipping, we set $\boldsymbol{h}_k = \boldsymbol{h}^\star$ and compare the proposed algorithm (17) referred to as "Proposed" and the conventional one (3) labeled "NLMS" in the Normalized Squared Error (NSE)

$$\mathrm{NSE}(k) := 10 \log_{10}((\gamma_k - \gamma^\star)^2/(\gamma^\star)^2)$$

averaged over 250 trials. In (3), the normalization constant is set as $w_\gamma = \|\boldsymbol{h}^\star\|_2^2$ according to [2,3].

Experiment with Gaussian input: The input signal x_k follows the i.i.d. Gaussian distribution $\mathcal{N}(0, 1)$, and v_k is the white Gaussian noise. The experiments are performed for different SNR conditions 15dB and 5dB and different initial estimates $\gamma_0 = 0$ and $\gamma_0 = 2$. In Fig. 2(a)(b), the conventional algorithm with $\gamma_0 = 2$ exhibits worse performance than one with $\gamma_0 = 0$ as the local minima trapping is caused by a relatively large initial estimate. Moreover, in Fig. 2(b) where SNR is 5dB, the performance of the conventional algorithms become unstable due to the local minima trapping as one of the trials is shown in Fig. 2(c). On the other hand, in all the experiments, the proposed algorithm shows a robust and better performance.

Experiment under double-talk: The input x_k is a speech signal in [36], and v_k consists of the white Gaussian noise and a speech in [36] summed in the double-talk occurring between the 5000 and the 10000th sample. The SNR between the echo signal y_k and the white Gaussian noise is 20dB, the SNR between y_k and the near-end speech is 5dB, and the initial estimate is set as $\gamma_0 = 0$. As shown in Fig. 2(d)(e), the conventional algorithm is trapped in local minima

[7] We denote $\boldsymbol{a}_{k,i} =: [a_{k,i,1}, \ldots, b_{k,i,r}]^t$ and $\boldsymbol{b}_{k,i} =: [b_{k,i,1}, \ldots, b_{k,i,r}]^t$.

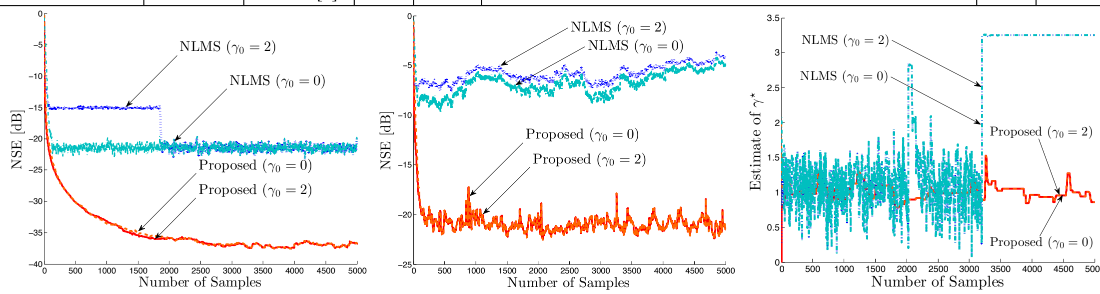
[8] We define an extended sign function $\overline{\mathrm{sgn}}(x, y) := \mathrm{sgn}(x)$ if $\mathrm{sgn}(x) \neq 0$, $\overline{\mathrm{sgn}}(x, y) := \mathrm{sgn}(y)$ otherwise, and define $\overline{\mathrm{sgn}}_r(\boldsymbol{x}, \boldsymbol{y}) := [\overline{\mathrm{sgn}}(x_1, y_1), \ldots, \overline{\mathrm{sgn}}(x_r, y_r)]^t \in \{-1, 0, 1\}^r$.

[9] The convex hull of $S \subset \mathbb{R}$ is the minimal convex set containing S and is denoted by $\mathrm{conv} S$. Note that the closedness of $\mathrm{conv} S_k^{(\varepsilon_k)}$ is guaranteed by the compactness of $S_k^{(\varepsilon_k)}$ [24, Theorem 1.4.3]. Remark that $S_k^{(\varepsilon_k)}$ is a single closed interval, i.e., $\mathrm{conv} S_k^{(\varepsilon_k)} = S_k^{(\varepsilon_k)}$, in most of our simulations.

[10] We define the ℓ_2 norm of $\boldsymbol{x} \in \mathbb{R}^n$ by $\|\boldsymbol{x}\|_2 := \sqrt{\boldsymbol{x}^t \boldsymbol{x}}$. For every $\boldsymbol{x} \in \mathbb{R}^n$ and nonempty closed convex set $C \subset \mathbb{R}^n$, the metric projection of $\boldsymbol{x}$ onto C is defined by $P_C(\boldsymbol{x}) := \arg\min_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|_2$, and the distance between $\boldsymbol{x}$ and C is defined by $d(\boldsymbol{x}, C) := \min_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|_2$.

Table 1: Parameter settings in the the experiments. The step sizes μ_γ and $\mu_{\boldsymbol{h}}$, $\tilde{\delta}_0$ and η are chosen in such a way that the initial convergence speed is same. In the proposed algorithm, $\varepsilon_{\rm mgn}^{(k)}$ is chosen to obtain the best steady-state behavior in our experiments. For the speech input, to cope with silent blocks in speech, $\varepsilon_{\rm mgn}^{(k)}$ of the proposed method is increased if $\|\boldsymbol{H}^\star \boldsymbol{x}_k\|_1$ is small. In addition to the previous case, $\varepsilon_{\rm mgn}^{(k)}$ of the "Proposed2" is increased if $\varepsilon_{\min}^{(k)} := \min_{\gamma\in[0,\gamma_{\max}]} \|\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k)\|_1$ becomes large, which implies unreliability of the sample due to the double-talk.

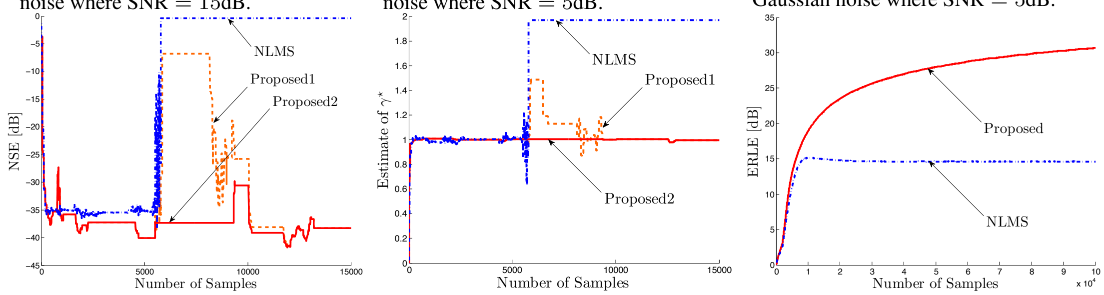
Situation	Input	Algorithm	μ_γ	r	$\varepsilon_{\rm mgn}^{(k)}$	$\mu_{\boldsymbol{h}}$	$\tilde{\delta}_0$	η
Hard-clipping Estimation	Gaussian	Proposed	1	50	8×10^{-4}	0	none	none
		NLMS [2]	0.6	none	none	0	none	none
	Speech	Proposed1	1	150	$\begin{cases}10^{-4}, & \text{if } \|\boldsymbol{H}^\star\boldsymbol{x}_k\|_1 > 3\times 10^{-4}\\ 10^4, & \text{otherwise.}\end{cases}$	0	none	none
		Proposed2	1	150	$\begin{cases}10^{-4}, & \text{if } \|\boldsymbol{H}^\star\boldsymbol{x}_k\|_1 > 3\times 10^{-4} \text{ and } \varepsilon_{\min}^{(k)} < 10^{-2}\\ 10^4, & \text{otherwise.}\end{cases}$	0	none	none
		NLMS [2]	0.5	none	none	0	none	none
Simultaneous Estimation	Gaussian	Proposed	0.1	150	5×10^{-3}	1.0	10^{-2}	0.998
		NLMS [5]	0.1	none	none	1.0	none	none



(a) NSE averaged in the experiments with Gaussian input and white Gaussian noise where SNR = 15dB.

(b) NSE averaged in the experiments with Gaussian input and white Gaussian noise where SNR = 5dB.

(c) Transition of γ_k in a trial of the experiments with Gaussian input and white Gaussian noise where SNR = 5dB.



(d) NSE averaged in the experiments under double-talk occurring between the 5000 and 10000th samples.

(e) Transition of γ_k in a trial of the experiments under double-talk occurring between the 5000 and 10000th samples.

(f) ERLE averaged in the experiments of the simultaneous adaptive learning of $\gamma^\star$ and $\boldsymbol{h}^\star$.

Fig. 2: Comparison of the proposed algorithm and the conventional algorithm.

due to the double-talk and is never updated even after the double-talk has ended. On the other hand, the proposed methods are free from the local minima trapping and have better steady-state behavior.

Next, to confirm the efficacy of the proposed simultaneous learning (17) and (18) referred to as "Proposed", we compare it with the conventional algorithm (3) and (4) labeled "NLMS" in the Echo Return Loss Enhancement (ERLE)

$$\mathrm{ERLE}(k) := 10\log_{10}\left(\sum_{j=k-\ell+1}^{k} y_j^2 \Big/ \sum_{j=k-\ell+1}^{k}(y_j - \hat{\boldsymbol{u}}_j^t \boldsymbol{h}_j)^2\right)$$

averaged over 250 trials with $\ell = 2000$. The input signal x_k follows the i.i.d. Gaussian distribution $\mathcal{N}(0,1)$, v_k is the white Gaussian noise with SNR = 15dB, and the initial estimates are set as $\gamma_0 = 2$ and $\boldsymbol{h}_0 = \boldsymbol{0}$. The cut-off value δ_k in (18) is designed according to [26, 27] with slight modifications:

$$\tilde{\delta}_k := \eta\tilde{\delta}_{k-1} + (1-\eta)\min\left\{\tilde{\delta}_{k-1},\, e_{k-1}^2(\gamma_{k-1}, \boldsymbol{h}_{k-1})/\|\hat{\boldsymbol{u}}_{k-1}\|_2^2\right\},$$
$$\delta_k = \|\hat{\boldsymbol{u}}_k\|_2\sqrt{\tilde{\delta}_k},$$

where $\eta\in(0,1)$ and $\tilde{\delta}_0 > 0$. In (3) and (4), the normalization constants are set as $w_\gamma = (\boldsymbol{h}_k^t \nabla\boldsymbol{\phi}_{\gamma_k}(\boldsymbol{x}_k))^2 + \omega_k \|\hat{\boldsymbol{u}}_k\|_2^2$ and $w_{\boldsymbol{h}} = \omega_k(\boldsymbol{h}_k^t \nabla\boldsymbol{\phi}_{\gamma_k}(\boldsymbol{x}_k))^2 + \|\hat{\boldsymbol{u}}_k\|_2^2$ with $\omega_k := 1/(\sqrt{N}\gamma_k)$ according to [5]. As shown in Fig. 2(f), the proposed method achieves the best steady-state behavior and improves about 16dB in the ERLE.

5. CONCLUDING REMARKS

For the nonlinear acoustic echo cancellation, we present an adaptive learning of the hard-clipping $\gamma^\star$ and the FIR system $\boldsymbol{h}^\star$ in (1) and (2). We solve the local minima trapping caused in the conventional algorithms [1–5] by introducing the feasible threshold set $S_k^{(\varepsilon_k)}$ in (6). To use the set in the estimation of the hard-clipping, we provide its explicit representation as a union of closed intervals $S_{k,i}^{(\varepsilon_k)}$ by using the piecewise linearity of $\boldsymbol{e}_k(\gamma, \boldsymbol{h}_k)$ w.r.t. γ. The proposed adaptive learning for the hard-clipping is derived by tracking the convex hull of the set $\mathrm{conv}S_k^{(\varepsilon_k)}$ based on the observations that $\mathrm{conv}S_k^{(\varepsilon_k)} = S_k^{(\varepsilon_k)}$ in most of our simulations. In the adaptive learning of $\boldsymbol{h}^\star$, we propose the use of the Huber loss function to measure $e_k(\gamma_k, \boldsymbol{h})$ for the robustness against the error in γ_k. Numerical examples show that the proposed algorithm is free from the local minima trapping and has an excellent steady-state behavior.

Future work includes an extension of the proposed algorithm to the framework [6, 7] where the echo path is modeled by the cascade of the hard-clipping, the power filter, and the FIR system.

6. REFERENCES

[1] B. Nollet and D. Jones, "Nonlinear echo cancellation for handsfree spcakerphones," in *Proc. NSIP*, 1997.

[2] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing*, vol. 80, no. 9, pp. 1747–1760, 2000.

[3] A. Stenger and W. Kellermann, "Nonlinear acoustic echo cancellation with fast converging memoryless preprocessor," in *Proc. IEEE ICASSP*, 2000, vol. 2, pp. 805–808.

[4] S. Shimauchi and Y. Haneda, "On initialization of cascaded nonlinear adaptive filter with hard clipping function," in *Proc. ASJ Autumn Meeting*, 2012, (in Japanese).

[5] S. Shimauchi, H. Ohmuro, and Y. Haneda, "Frequency domain implementation of cascaded nonlinear adaptive filter with hard clipping function," in *Proc. IEICE SIP Symposium*, 2012, (in Japanese).

[6] M. I. Mossi, C. Yemdji, N. W. D. Evans, and C. Beaugeant, "A cascaded non-linear acoustic echo canceller combining power filtering and clipping compensation," Tech. Rep. RR-11-258, Eurecom, Aug. 2011.

[7] M. I. Mossi, C. Yemdji, N. Evans, C. Beaugeant, F. M. Plante, and F. Marfouq, "Dual amplifier and loudspeaker compensation using fast convergent and cascaded approaches to nonlinear acoustic echo cancellation," in *Proc. IEEE ICASSP*, 2012, pp. 41–44.

[8] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proc. IEEE ICASSP*, 2005, vol. 3, pp. 105–108.

[9] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Autom. Control*, vol. 12, no. 3, pp. 282–287, Jun. 1967.

[10] L. M. Bregman, "The method of successive projections for finding a common point of convex sets," *Soviet Math. Dokl.*, vol. 6, pp. 688 – 692, 1965.

[11] L. G. Gubin, B. T. Polyak, and E. V. Raik, "The method of projections for finding the common point of convex sets," *USSR Comput. Math. and Phys.*, vol. 7, no. 6, pp. 1 – 24, 1967.

[12] H. Bauschke and J. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Rev.*, vol. 38, no. 3, pp. 367–426, 1996.

[13] H. Stark and Y. Yang, *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics*, Wiley, New York, 1998.

[14] I. Yamada, "Adaptive projected subgradient method: A unified view for projection based adaptive algorithms," *J. IEICE*, vol. 86, no. 8, pp. 654–658, Aug. 2003, (in Japanese).

[15] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numer. Funct. Anal. and Optim.*, vol. 25, no. 7&8, pp. 593–617, 2005.

[16] K. Slavakis, I. Yamada, and N. Ogura, "The adaptive projected subgradient method over the fixed point set of strongly attracting nonexpansive mappings," *Numer. Funct. Anal. and Optim.*, vol. 27, no. 7-8, pp. 905–930, 2006.

[17] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Stat.*, vol. 35, no. 1, pp. 73–101, 1964.

[18] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," in *Proc. IEEE ICASSP*, Mar. 1999, pp. 877–880.

[19] O. Hoshuyama and A. Sugiyama, "Nonlinear echo cancellation based on spectral shaping," in *Speech and Audio Processing in Adverse Environments*, E. Hänsler and G. Schmidt, Eds., Signals and Communication Technology, pp. 267–283. Springer, 2008.

[20] D. Meng, Z. Xu, L. Zhang, and J. Zhao, "A cyclic weighted median method for L_1 low-rank matrix factorization with missing entries," in *Proc. AAAI*, 2013.

[21] D. R. K. Brownrigg, "The weighted median filter," *Commun. ACM*, vol. 27, no. 8, pp. 807–818, Aug. 1984.

[22] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, "Weighted median filters: a tutorial," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 43, no. 3, pp. 157–192, 1996.

[23] A. Rauh and G. R. Arce, "Optimal pivot selection in fast weighted median search," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4108–4117, 2012.

[24] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms: Part 1: Fundamentals*, vol. 1, Springer, 1996.

[25] P. Petrus, "Robust Huber adaptive filter," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 1129–1133, 1999.

[26] L. R. Vega, H. Rey, J. Benesty, and S. Tressens, "A new robust variable step-size NLMS algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1878–1893, 2008.

[27] T. Yamamoto, M. Yamagishi, and I. Yamada, "Adaptive proximal forward-backward splitting for sparse system identification under impulsive noise," in *Proc. EUSIPCO*, Aug. 2012, pp. 2620–2624.

[28] Y. Murakami, M. Yamagishi, M. Yukawa, and I. Yamada, "A sparse adaptive filtering using time-varying soft-thresholding techniques," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 3734–3737.

[29] M. Yamagishi, M. Yukawa, and I. Yamada, "Sparse system identification by exponentially weighted adaptive parallel projection and generalized soft-thresholding," in *Proc. APSIPA ASC*, Dec. 2010, pp. 367–370.

[30] M. Yukawa, "Multikernel adaptive filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4672–4682, Sep. 2012.

[31] M. Yukawa, Y. Tawara, M. Yamagishi, and I. Yamada, "Sparsity-aware adaptive filters based on ℓ_p-norm inspired soft-thresholding technique," in *Proc. IEEE ISCAS*, May 2012, pp. 2749–2752.

[32] H. Kuroda, S. Ono, M. Yamagishi, and I. Yamada, "Exploiting group sparsity in nonlinear acoustic echo cancellation by adaptive proximal forward-backward splitting," *IEICE Trans.*, vol. E96-A, no. 10, pp. 1918–1927, Oct. 2013.

[33] S. Ono, M. Yamagishi, and I. Yamada, "A sparse system identification by using adaptively-weighted total variation via a primal-dual splitting approach," in *Proc. IEEE ICASSP*, May 2013.

[34] I. Yamada, S. Gandy, and M. Yamagishi, "Sparsity-aware adaptive filtering based on a Douglas-Rachford splitting," in *Proc. EUSIPCO*, Sep. 2011, pp. 1929–1933.

[35] M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proc. 16th Int. Conf. DSP*, 2009, pp. 1–5.

[36] P. Kabal, "TSP speech database," Tech. Rep., Department of Electrical & Computer Engineering, McGill University, Montreal, Quebec, Canada, 2002.