ON DETECTION OF DOUBLE TALK AND CHANGES IN THE ECHO PATH USING A MARKOV MODULATED CHANNEL MODEL

Catharina Carlemalm

Andrew Logothetis *

S3-Automatic Control, Royal Institute of Technology, S-100 44 Stockholm, Sweden, email: cath@s3.kth.se Department of Electrical and Electronic Engineering, University of Melbourne, Australia, email: a.logothetis@ee.mu.oz.au

ABSTRACT

This paper addresses the problem of detection and discrimination of the two phenomena double talk and echo path change in a telephone channel. It is of uttermost importance to quickly detect a change in the echo path while not confusing it with double talk, since the echo canceller should react differently whether an echo path change or double talk has occurred. In this paper, novel algorithms of low complexity are proposed. The system is described with a Markov modulated finite impulse response (FIR) filter. Depending on whether double talk or an echo path change occurs different parameters in the channel model change abruptly. Based on model assumptions, maximum likelihood (ML) parameter estimates of the communication channel are obtained via recursive (off-line) or iterative (on-line) methods using the expectation maximization (EM) algorithm. This enables us to use a Hidden Markov Model (HMM) state estimator to yield the minimum probability of error in identifying the state of the communication channel, *i. e.* the possible presence of double talk and/or echo path change. The proposed algorithms are experimentally verified using a real speech signal and impulse responses created from measured impulse responses from real hybrids.

1. INTRODUCTION

The objective of this paper is to propose a new method for detection and discrimination of double talk and changes in the echo path. This problem has previously been studied in [2] where a likelihood based detection scheme is suggested which compares a global channel model with a local one, both estimated with the Recursive Least Squares (RLS) algorithm.

Commonly utilized methods for reducing echo in the telephone network are echo cancellers. The idea of the echo canceller is to use a finite impulse response (FIR) filter to approximate the transfer function of the echo path and let the filter coefficients be adjusted based on the calculated prediction error [1]. This simple solution to the problem of echo control is unfortunately not applicable due to the existence of double talk, that is, both subscribers talk simultaneously. This phenomenon also gives rise to an abrupt increase of the prediction error. One must avoid making large corrections during double talk of the echo path in a doomed-to-failure attempt to cancel the echo. Thus, the adaptation rate should be decreased during double talk.

The conclusions to be made are that merely measuring the prediction error will not discriminate between double talk and echo path changes and that the echo canceller must react differently whether double talk or a change in the echo path has occurred. Furthermore, the detection and discrimination algorithm must be fast in order to prevent the FIR filter in the echo canceller from being misadjusted.

The key contribution of this paper is to use a Markov modulated FIR filter to describe the system. The theory of Markov modulated FIR filters is not new, see [3], but the application of the theory to this problem is genuinely new. By stating hypotheses corresponding to the different phenomena and by using the Expectation Maximization (EM) algorithm, an off-line detection scheme is formulated. Furthermore, an on-line algorithm is suggested. The on-line detection scheme is experimentally verified using real data.

This paper is organized as follows. In Section 2 the problem is stated. Furthermore, the model and the notation used in the following sections are introduced and assumptions are made. Section 3 the proposed algorithms are derived. Next, in Section 4 some experimental results are presented when real data is used. Finally, Section 5 gives some concluding remarks.

2. PROBLEM DESCRIPTION

2.1. Signal Model

The system is described with a Markov modulated finite impulse response (FIR) filter. The output of the filter is given by

$$y_k = c_0(s_k)u_k + \ldots + c_{n-1}(s_k)u_{k-n+1} + \sqrt{d(r_k)v_k}, \quad (1)$$

$$k = 1, 2, \ldots$$

where $\{v_k\}$ is a zero mean white Gaussian process with variance 1, $\{u_k\}$ is the *known* input signal, $\{s_k\}$ and $\{r_k\}$ are two independent *N*-state, discrete time, homogeneous first order Markov processes and finally $c_j(i)$, $i = 1, \ldots, N$, $j = 1, \ldots, n-1$ and d(i), $i = 1, \ldots, N$ are constants. The transition probability matrix of $\{s_k\}$ is denoted by $A^{(s)} = \begin{pmatrix} a_{ij}^{(s)} \\ a_{ij}^{(s)} \end{pmatrix}$, where $a_{ij}^{(s)} = \Pr(s_{k+1} = j | s_k = i)$. Furthermore, $a_{ij}^{(s)} \ge 0$ and $\sum_{j=1}^{2} a_{ij}^{(s)} = 1$ for each *i*. Similar notation and conditions hold for $\{r_k\}$. To facilitate the reading the following notations are

To facilitate the reading the following notations are used $C(i) = (c_0(i), c_1(i), \ldots, c_{n-1}(i))$ for $i = 1, \ldots, N$, $Y_T = (y_1, y_2, \ldots, y_T)$ denotes the measurement sequence and $\theta = (C(i), d(i), A^{(s)}, A^{(r)}; i = 1, \ldots, N)$ denotes the model parameter vector. Finally, let θ_0 denote the true model parameters.

Remark: A more accurate and realistic model for describing the communication channel (instead of Eq. (1)) is the following

$$y_k = \sum_{m=0}^{n-1} c_m(s_{k-m}) u_{k-m} + \sqrt{d(r_k)} v_k$$
(2)

Eq. (1) is used instead of (2) for the following reasons: 1) Computationally less expensive when computing estimates of the channel state. 2) If it is not crucial to detect the possible abrupt change within a channel length, n.

^{*}Supported by ATERB and CSSIP.

2.2. Statement of Hypotheses

The state s_k (or r_k) at time k is an element of the set $Q = \{1, 2\}$. s_k denotes the current state of the echo path, while r_k denotes the presence or non-existence of double talk.

Consider the system (1) at time k under the following different hypotheses concerning the states s_k and r_k .

$$\begin{array}{rrrr} H_0: & s_k = 1 & \text{and} & r_k = 1 \\ H_1: & s_k = 1 & \text{and} & r_k = 2 \\ H_2: & s_k = 2 & \text{and} & r_k = 1 \\ H_3: & s_k = 2 & \text{and} & r_k = 2 \end{array}$$

The situation H_1 occurs when double talk (DT) but no echo path change (EPC) appears at time k, the situation H_2 when an EPC but no DT happens at time k, hypothesis H_3 when both an EPC and DT appear at time k and finally the null hypothesis when neither EPC nor DT has occurred at time k.

To sum up, an EPC causes abrupt changes in the parameters c_j while DT gives rise to an abrupt change in the noise variance d. Similar hypotheses were used in [2].

2.3. Objectives

The objective of this paper is to suggest two different detection schemes, one off-line and one on-line. These are briefly described below.

- Off-line channel identification and change detection Given all the data y_1, \ldots, y_T , it is desired to perform the following two steps:
 - 1. *Parameter Estimation:* Compute the maximum likelihood (ML) parameter estimate of the model

$$\theta^{ML} \stackrel{\Delta}{=} \arg\max_{a} f(Y_T|\theta) \tag{3}$$

where $f(\cdot)$ is the probability density function.

2. Detection (State Estimation): Decide which hypothesis holds, *i. e.* the possible presence of double talk and/or change in the echo path with a minimum probability of error based on the best (ML) estimate of the channel. In the Bayesian framework this is achieved by

$$\hat{s}_k = \arg \max_{s_k} f\left(s_k \mid Y_T, \theta^{ML}\right) \tag{4}$$

$$\hat{r}_k = \arg \max_{r_k} f(r_k \mid Y_T, \theta^{ML})$$
(5)

- On-line channel identification and change detection Given the observations y_1, \ldots, y_l up to time l, it is desired to perform the following two steps:
 - 1. Parameter Estimation: Compute the maximum likelihood (ML) parameter estimate of the model

$$\Theta^{(l)} \stackrel{\Delta}{=} \arg\max_{\theta} f(Y_l|\theta) \tag{6}$$

where $\theta^{(l)}$ is the ML parameter estimate based on the first l data points.

2. Detection (State Estimation): Decide which hypothesis holds, *i. e.* the possible presence of double talk and/or change in the echo path, by computing

$$\hat{s}_{l} = \arg \max_{s_{l}} f\left(s_{l} \mid Y_{l}, \theta^{(l-1)}\right)$$
(7)

$$\hat{r}_{l} = \arg \max_{r_{l}} f\left(r_{l} \mid Y_{l}, \theta^{(l-1)}\right)$$
(8)

Remark: A theoretical analysis of the effects of initial estimates is beyond the scope of this paper. Furthermore, the question relating to identifiability is very difficult and not studied in this paper.

3. ANALYSIS

The expectation maximization (EM) algorithm suggested in [6] is used to obtain θ^{ML} . Also, as a by-product of the E-step, the maximum a posteriori (MAP) estimates of the states s_k and r_k are obtained.

It is shown in [7] that under mild regularity conditions, the sequence $\{\theta^{(l)}\}$ of the EM algorithm converges to a stationary value of the likelihood function. *EM algorithm:*

- 0. Determine the initial estimate $\theta^{(1)}$.
- 1. (E-step) Evaluate

$$Q\left(\theta, \theta^{(l)}\right) \stackrel{\Delta}{=} \mathrm{E}\left\{\ln f(Y_T, S_T, R_T \mid U_T, \theta) | U_T, Y_T, \theta^{(l)}\right\}$$
(9)

2. (M-step) Compute

$$\theta^{(l+1)} = \arg\max_{a} Q\left(\theta, \theta^{(l)}\right) \tag{10}$$

3. l := l + 1. Iterate steps 1–3 until $\|\theta^{(l+1)} - \theta^{(l)}\| < \varepsilon$, where ε is some specified constant.

3.1. Off-line Algorithm

First, the *l*:th iteration of the E-step in the above mentioned EM algorithm is performed. For notational convenience we drop the dependence of the equations on l and U_T .

$$\ln f(Y_T, S_T, R_T) = \sum_{k=1}^{T} \ln f(y_k | s_k, r_k) + \ln f(r_1) + \ln f(s_1) + \sum_{k=2}^{T} [\ln f(s_k | s_{k-1}) + \ln f(r_k | r_{k-1})]$$
(11)

The input signals, u_{-n}, \ldots, u_0 are assumed to be known.

Taking the expected value of (11) and making the approximation that the first n-1 terms can be ignored $(n \ll T)$, results in

$$Q\left(\theta, \theta^{(l)}\right) = \sum_{k=n}^{T} \sum_{i,j=1}^{2} \ln\left(a_{ij}^{(s)}\right) \gamma_{k}^{(s)}(i,j) + \sum_{k=n}^{T} \sum_{i,j=1}^{2} \ln\left(a_{ij}^{(r)}\right) \gamma_{k}^{(r)}(i,j) - \frac{T-n+1}{2} \ln(2\pi) - \frac{1}{2} \sum_{k=n}^{T} \sum_{i=1}^{2} \ln(d_{i}) \gamma_{k}^{(r)}(i) - \frac{1}{2} \sum_{k=n}^{T} \sum_{i=1}^{2} d_{i}^{-1} \times \sum_{j=1}^{2} \left(y_{k} - \sum_{m=0}^{n-1} c_{m}(j) u_{k-m}\right)^{2} \gamma_{k}^{(s)}(j) \gamma_{k}^{(r)}(i)$$
(12)

where

$$\gamma_k^{(s)}(i) \stackrel{\Delta}{=} \mathbf{P}\left(s_k = i | Y_T, \theta^{(l)}\right), \ i \in \{1, 2\}$$
(13)

and

$$\gamma_k^{(s)}(i,j) \stackrel{\Delta}{=} \mathbf{P}\left(s_k = j, s_{k-1} = i | Y_T, \theta^{(l)}\right), \ i, j \in \{1,2\}$$
(14)

Similar notations hold for the Markov chain r.

Maximizing (12) with respect to θ , *i. e.* performing the M-step gives

I. Update of the transition probability matrix, A.

$$a_{ij}^{(p)} = \frac{\sum_{k=n+1}^{T} \gamma_{k-1}^{(p)}(i,j)}{\sum_{k=n}^{T-1} \gamma_{k}^{(p)}(i)},$$
(15)

II. Update of the noise variance, d.

$$d_{i} = \frac{\sum_{k=n}^{T} \sum_{j=1}^{2} \left(y_{k} - \sum_{m=0}^{n-1} c_{m}(j) u_{k-m} \right)^{2} \gamma_{k}^{(s)}(j) \gamma_{k}^{(r)}(i)}{\sum_{k=n}^{T} \gamma_{k}^{(r)}(i)}$$
(16)

for $p = r, s, i, j \in \{1, 2\}$.

III. Update of the coefficients, C(j).

The elements $c_m(j)$ of C(j), $m = 0, \ldots, n-1, j \in \{1, 2\}$ are given by T 2

$$\sum_{k=n}^{T} \sum_{i=1}^{2} d_{i}^{-1} y_{k} \gamma_{k}^{(s)}(j') \gamma_{k}^{(r)}(i) u_{k-m'}$$

$$= \sum_{k=n}^{T} \sum_{i=1}^{2} d_{i}^{-1} \sum_{m=0}^{n-1} c_{m}(j') u_{k-m} \gamma_{k}^{(s)}(j') \gamma_{k}^{(r)}(i) u_{k-m'}$$
(17)

for $m' = 0, \ldots, n - 1, j' = 1, 2$.

Finally, the forward probabilities for $k = n, \ldots, T$, are given bv

$$\alpha_{k}(j_{r}, j_{s}) \stackrel{\Delta}{=} f(Y_{k}, s_{k} = j_{s}, r_{k} = j_{r}|\theta)$$

$$= \sum_{i_{s}=1}^{2} \sum_{i_{r}=1}^{2} f(y_{k}|s_{k} = j_{s}, r_{k} = j_{r}) a_{i_{s}j_{s}}^{(s)} a_{i_{r}j_{r}}^{(r)} \alpha_{k-1}(i_{r}, i_{s})$$
(18)

$$\alpha_m(j_r, j_s) = \frac{1}{4}, \text{ for } m = 1, \dots, n-1$$
 (19)

and the backwards probabilities by

$$\beta_{k}(i_{r}, i_{s}) \stackrel{\Delta}{=} f\left(Y_{k+1}^{T} | s_{k} = i_{s}, r_{k} = i_{r}, Y_{k}\right)$$

$$= \sum_{j_{s}=1}^{2} \sum_{j_{r}=1}^{2} \beta_{k+1}(j_{r}, j_{s}) f(Y_{k+1} | s_{k+1} = j_{s}, r_{k+1} = j_{r})$$

$$\times a_{i_{s}, j_{s}}^{(s)} a_{i_{r}, j_{r}}^{(r)}$$
(20)

for $i_r, i_s, j_r, j_s \in \{1, 2\}$. Initialisation of the backward prob-

abilities are $\beta_{T+1}(i_r, i_s) = 1$, $i_r, i_s \in \{1, 2\}$. Using the relations above, the following results are achieved

$$\gamma_k^{(r,s)}(i_r, i_s) = \frac{\alpha_k(i_r, i_s)\beta_k(i_r, i_s)}{\sum_{i_s=1}^2 \sum_{i_r=1}^2 \alpha_k(i_r, i_s)\beta_k(i_r, i_s)}$$
(21)

and

$$\gamma_k^{(s)}(i_s) = \sum_{i_r=1}^2 \gamma_k^{(r,s)}(i_r, i_s)$$
(22)

for $i_r, i_s \in \{1, 2\}$.

IV. Channel state estimation.

The minimum probability of error estimate of the state of the channel is achieved by

$$\hat{s}_{k}^{MAP} = \arg \max_{s_{k}} f\left(s_{k} \mid Y_{T}, \theta^{(l)}\right)$$
$$\hat{r}_{k}^{MAP} = \arg \max_{r_{k}} f\left(r_{k} \mid Y_{T}, \theta^{(l)}\right)$$
(23)

The memory and computational requirements of the offline algorithm are shown Table 1.

	Off-line per pass		On-line	
	Memory	Comp.	Memory	Comp.
γ_k	$2^{2}T$	$O(2^2T)$	2^{2}	$O(2^2)$
A	2×2^2	$O(2^2T)$	2×2^2	$O(2^2)$
C	2n	$O(n^3T)$	2n	O(n)
d	2	O(nT)	2	O(n)

Table 1. Memory and computational requirements of the algorithms.

3.2. On-line Stochastic Gradient Algorithm

The on-line stochastic gradient algorithm studied in [4] is applied on the posed problem. That is, the parameters are updated according to

$$\theta^{(k)} = \theta^{(k-1)} + \frac{1-\rho}{1-\rho^{k-1}}K(k)I_k$$
(24)

where

$$I_{k} = \rho I_{k-1} + S\left(\theta^{(k-1)}, k\right), \ I_{1} = S\left(\theta^{(k-1)}, k\right), \ I_{0} = 0$$
(25)

where $K(k) = 1/k^m$ and the incremental score vector is defined by $S\left(\theta^{(k-1)},k\right) \stackrel{\Delta}{=}$

 $\left. \frac{\partial \mathbb{E}\left\{ \ln f(y_k, s_k, r_k | Y_{k-1}, S_{k-1}, R_{k-1}, \theta^{(k-1)} \right\}}{\partial \theta} \right|_{\theta = \theta^{(k-1)}}$ (26)

I. Update of the transition probabilities, A.

Details on the recursive update of transition probabilities based on a differential geometric approach is found in [5]. Due to space limitation we omit details.

II. Update of the noise variance, d.

The element of the incremental score vector corresponding to $d(r_k = i_r)$, $i \in \{1, 2\}$, is given by 9

$$d^{-2}(r_k = i_r, k-1) \sum_{i_s=1}^{2} \left(y_k - \sum_{l=0}^{n-1} c_l(s_k = i_s, k-1) u_{t-l} \right)^2 \times \alpha_k(i_r, i_s) - d^{-1}(r_k = i_r, k-1) \alpha_k^{(r)}(i_r),$$
(27)

where $d(i_r, k-1)$ is the estimate at time k-1 and $\alpha_k^{(r)}(i_r) \equiv$ $P(Y_k, r_k = i_r).$

III. Update of the coefficients, C(j).

The element of the incremental score vector corresponding to $c(s_k = i_s), i_s \in \{1, 2\}$, is given by

$$\sum_{i_r=1}^{2} \frac{y_k - \sum_{l=0}^{n-1} c_l(s_k = i_s, k-1) u_{k-l}}{d(r_k = i_r, k-1)} u_{k-n} \alpha_k(i_r, i_s).$$
(28)

In Table 1 the memory and computational requirements of the on-line algorithm are outlined.

4. EXPERIMENTAL RESULTS

In this section the on-line method which has been proposed in this paper for detection and discrimination of double talk and echo path changes is evaluated by simulations using a real speech signal of length 4000 samples as input signal. It is depicted in Fig. 1. The sampling frequency is 8000 Hz. To generate the different hypotheses, the input signal is segmented. A change in the echo path is simulated by filtering the segments through different impulse responses. The impulse responses used are shown in Fig. 1. They are constructed from measured impulse responses from real hybrids using the most active part of the measured impulse response. Motivation for this approach can be found in [8]. Double talk is simulated by filtering the segments through the same impulse response and then adding speech to one of the segments. The parameters used in the simulations are given in Table 2. The transition probabilities and the noise levels are not estimated and are assumed to be known a priori.

In Fig. 2 the root mean square estimation errors of the channel coefficients for the on-line scheme are shown. Fig. 3 depicts the true and estimated state of the communication channel, *i. e.* the possible presence of double talk and/or $\frac{1}{2}$ echo path change. As can be seen from these figures the on-line method succeeds in detecting the phenomenon double talk. The echo path change that occurs simultaneously as double talk is not detected. This is not disastrous since double talk is the phenomenon that is most crucial to alarm for, since the adaptation of the echo canceller should then be stopped. Though, all other echo path changes are successfully detected. The abrupt change in the error while estimating the coefficients of the first channel at approximately sample time k = 1000 is due to two main reasons: 1) A short delay in detecting the echo path change. 2) Using a zero forgetting factor ($\rho = 0$). Thus, increasing the forgetting factor and/or using more data points, the algorithm will slowly adapt to the true parameters.

For the examples described above, the on-line detection and discrimination algorithm suggested in this paper seems reliable. Furthermore, it has been observed, in simulations studies not presented here, that the proposed on-line algorithm performs extremely well for known statistics.

Number of samples	T = 4000
Number of taps	n = 50
Values of d (known)	d(1) = 0.1, d(2) = 5000
Transition probabilities of s	$a_{11}^{(s)}=a_{22}^{(s)}=0.999,\ a_{12}^{(s)}=a_{21}^{(s)}=0.001$
Transition probabilities of r	$a_{11}^{(r)}=0.999, a_{22}^{(r)}=0.997,\ a_{12}^{(r)}=0.001, a_{21}^{(r)}=0.003$

Table 2. Parameters used in the simulations.



Figure 1. Input signal and impulse responses.

5. CONCLUSIONS

In this paper two algorithms for detection and discrimination of double talk and changes in the echo path were proposed. Their complexities were shown to be of order O(n), where *n* denotes the channel length. The communication channel was modelled with a Markov modulated FIR filter. Therefore, abrupt changes in the channel characteristics could be effectively included. Using the expectation maximization algorithm, the maximum likelihood estimates of the channel parameters were determined. By identifying hypotheses corresponding to the two phenomena, two algorithms could be formulated; one on-line and one off-line.



Figure 2. Root mean square estimation error of the channel coefficients for the on-line scheme.



Figure 3. True and estimated state of the communication channel.

Simulations showed that the performance of the on-line algorithm was satisfying when real speech was used as input signal and impulse responses created from measured impulse responses from real hybrids. Due to space limitations, simulations using the off-line scheme had to be omitted.

REFERENCES

- N.A.M. Verhoeckx, H.C. van den Elzen, F.A.M. Snijders and P.J. van Gerwen, "Digital Echo Cancellation for Baseband Data Transmission", *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol. ASSP-27, no.6, pp.768-781, 1979.
- [2] C. Carlemalm, F. Gustafsson and B. Wahlberg, "On the problem of detection and discrimination of double talk and change in the echo path", Proc. ICASSP'96, Atlanta, Georgia, USA, pp. 2742–2745, 1996.
- [3] S. Dey, V. Krishnamurthy and T. Salmon-Legagneur, "Estimation of Markov-modulated time-series via EM algorithm", *IEEE Signal Proc. Letters*, vol. 1, no. 10, pp. 153-155, 1994.
- [4] V. Krishnamurthy and A. Logothetis, "Iterative and recursive estimators for hidden Markov errors-invariables models", *IEEE Trans. on Signal Proc.*, vol. 44, no. 3, pp. 629–639, 1996.
- [5] I. B. Collings, V. Krishnamurthy, J. B. Moore, "Online identification of hidden Markov models via recursive prediction error techniques", *IEEE Trans. Signal Proc.*, Vol. 42, No. 12, pp. 3535-3539, Dec.1994.
- [6] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", J. Roy. Stat. Soc., vol. 39, pp. 1–38, 1977.
- [7] C. F. J. Wu, "On the convergence properties of the EM algorithm", *The Annals of Statistics*, Vol. 11, No. 1, pp. 95–103.
- [8] C. Carlemalm, "A novel low complexity method for detection and discrimination of double talk and echo path changes", *submitted to ECSAP*'97, Prague, Czech Republic, 1997.