Telephony acoustics at Mitel

Philippe Moquin
Mitel Corp., Kanata, ON

1. Introduction

This is an overview of Mitel's acoustical research and development in the desktop terminals group. Telephones and operator consoles are the primary products designed by the group.

2. Handset performance

Basic performance of the handset is strongly standardised however there are a few areas that still affect the sound quality: wind noise and receiver real ear performance.

The microphone wind noise during explosive consonants proved to be a significant problem on a new handset and due to time constraints a co-operative research project with Dr. Stinson of NRC was set up. Results from the research yield an interim design that significantly improved the performance. Further research provided us an opportunity to reduce the wind noise to a level that is among the lowest in the industry. An accurate analytical model was also developed. Having a good physical understanding of the noise generation mechanism proved to be very valuable in the redesign.

In standardised tests the artificial ears used accurately replicate a handset perfectly sealed to the handset. Unfortunately few people hold a handset that tightly to their ear. The air leak between the ear and the handset creates a significant loss of low frequency sensitivity. The majority of existing receivers are designed for the sealed ear condition. However, we anticipate that, with the advent of high quality personal music systems, clients will soon demand better performing receivers.

3. Speakerphones

All of Mitel's telephones are made in Kanata and feature high quality speakerphones. Three years ago Mitel had no acoustical modelling capability. It was decided that a university research project could provide that capability. The project has started and we are sponsoring a doctoral candidate under Dr. Laville's direction at l'école de Technologie Supérieure.

We are looking to model the transducer frequency response in the plastic housings and microphone to speaker acoustic separation. The approach taken is rather novel as it integrates analytical, empirical and numerical methods to provide a computationally efficient and reasonably accurate model.

High quality speakerphones today necessarily imply full duplex operation. This unfortunately brings up the "barrel effect" where the received signal is retransmitted to the far end talker providing an annoying echo.

The present preferred solution is to use an adaptive FIR filter that mimics the room impulse response. This has some significant draw backs.

Firstly, there is the computational requirements to model a reasonably live room. There is much work being done to develop more efficient algorithms.

Secondly, there is the finite performance possible with large FIR filters. New approaches to solve the problem are needed that do not have the performance limitations inherent in conventional NLMS FIR algorithms. This requires a good understanding of both telephony and acoustics of rooms.

Thirdly, this in no way addresses the microphone to talker distance. Methods for reducing the effective talker to microphone have to be developed. We feel that microphone arrays such as those described by Drs. Ryan and Stinson will be useful.

Fourthly, these algorithms assume absolute linearity. Non-linearity due to button rattle and other distortions can be controlled by physical design but some non-linearity's would be most useful.

4. Speech recognition

Mitel is not proposing to do significant research in speech recognition since we believe that we will be able to licence this type of technology. However, the acoustical design of telephones will have to accommodate speech recognition. Obviously the challenge is greater in the speakerphone mode. The primary problem is that of providing the best signal to noise for the talker signal. Again the belief is that directional microphones or arrays will provide significant performance enhancement.

5. Voice over IP

The problem with transmitting voice or audio visual stream over computer network is that you don't have a dedicated channel as one has in a telephone network. Packets are sent out serially as in the telephone network but the time to get to the other end is not known. The packet may never get there, it may arrive late and even if they do get the in order the timing between one packet and the next may vary significantly.

Presently, there are only theories as to the best way to deal with these problems. Large buffers can be used but this means long time delays which can inhibit communication. Lost or late packets can be replaced by silence or noise but this may not be acceptable to users of wired telephones while it works fine in the wireless world. Alternative coding schemes with built in redundancy could help but then one is locked into one vendor's solution.

We believe that using conventional technology with a good understanding of speech perception creative solutions providing toll quality voice over IP networks is achievable. This area obviously requires significant research effort.

On the use of the FNTF algorithm in subband acoustic echo cancellation

Benoît Champagne and François Caron
INRS-Télécommunications, Université du Québec, 16 place du Commerce
Verdun, Québec, Canada, H3E 1H6; champagne@inrs-telecom.uquebec.ca

1. Introduction

Acoustic echo cancellation (AEC) is an effective approach for the control of acoustic echoes generated by hands-free terminals [1]. Even though RLS adaptive filtering achieves faster convergence, the NLMS algorithm is used in most AEC applications because of its low complexity of $2N$ multiplies per iteration (mpi), where N is the number of filter taps. The identification of the long echo paths found in teleconference applications (e.g. $N = 1000$ at 8kHz sampling rate) further requires the use of subband processing to bring down the computational complexity within acceptable limits [2].

Recently, the fast Newton transversal filter (FNTF) algorithms have been proposed in an attempt to bridge the performance gap between the NLMS and the fast RLS (FRLS) algorithms [3]. FNTF models the excitation signal as an AR(M) process, where $0 \leq M \leq N$, and achieves a complexity of $2N + 12M$. In [4], FNTF is shown to be an attractive candidate for AEC applications in the mobile context (short filters, $N = 250$ at 8kHz sampling) since significant improvements in convergence speed over NLMS can be obtained with small values of M. This conclusion does not hold for teleconference applications, where FNTF may lead to a loss of performance.

The investigation in [4] is limited to the use of a single (i.e. full-band) transversal adaptive filter. In the case of long impulse responses, FNTF might benefit from a subband implementation because of the reduced adaptive filter length in subbands, as a result of downsampling. In this work, we investigate the performance of FNTF in a subband AEC structure, with emphasis on the identification of long echo paths typical of teleconference applications.

2. The FNTF algorithm

Application of adaptive system identification to AEC is illustrated in Fig. 1. The unknown system $\mathcal{H}$ consists of the loudspeaker, acoustic medium and microphone. The system input is the far-end signal $u(n)$, where $n \in \{1, 2, ...\}$ is the discrete-time, and the system output is the microphone signal $d(n)$, which contains additive noise and possibly local speech. The unknown system is modeled by an adaptive transversal FIR filter operating on $u(n)$. The time-varying coefficients of the FIR filter are denoted by $h_k(n)$, $k = 0, 1, ..., N-1$, and the filter output is computed as $\hat{d}(n) = \mathbf{h}(n)^T \mathbf{u}(n)$, where $\mathbf{h}(n) = [h_0(n), ..., h_{N-1}(n)]^T$ and $\mathbf{u}(n) = [u(n), u(n-1), ..., u(n-N+1)]^T$. The filter weight vector is recursively adjusted in real-time so as to minimize the power of the error signal, defined as $e(n) = d(n) - \hat{d}(n)$. Practical operation of the adaptive filter requires the use of a double-talk detector (not considered in this study).

The FNTF algorithms belong to a modified class of stochastic Newton (SN) adaptive algorithms:

$$\mathbf{c}_N(n) = -\lambda^{-1} R_N^{-1}(n-1)\mathbf{u}(n), \quad \gamma(n) = 1 - \mathbf{c}_N^T(n)\mathbf{u}(n) \quad (1)$$

$$e(n) = d(n) - \mathbf{h}^T(n)\mathbf{u}(n), \quad \epsilon(n) = e(n)/\gamma(n) \quad (2)$$

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \epsilon(n)\mathbf{c}_N(n) \quad (3)$$

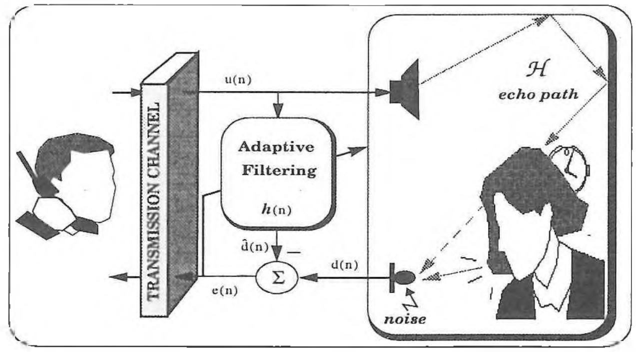

Figure 1: Adaptive identification applied to AEC

where $R_N(n)$ is an estimate of the data covariance matrix, $0 < \lambda \leq 1$ is a forgetting factor, $\mathbf{c}_N(n)$ is a generalized dual Kalman gain, $e(n)$ is the *a priori* estimation error, $\epsilon(n)$ is the *a posteriori* error, and $\gamma(n)$ is a conversion factor. Both NLMS and FRLS can be obtained from (1)-(3) with a proper choice of $R_N(n)$ in (1). In [3], an additional AR(M) assumption on $u(n)$, where $0 \leq M \leq N$, is exploited to derive time-order recursions for the extension of an $(M+1)$th order covariance matrix into the desired N-order one, i.e. $R_N(n)$. Upon substitution of these extension formulae in (1), three distinct FNTF versions are obtained with complexity $2N + O(M)$.

In the context of AEC, practical considerations point to the use of FNTF Version 1 [4]. The latter, used in our work, is summarized below: (a) Using a FRLS forward predictor of order M applied to $u(n)$, update the forward predictor weight vector $\mathbf{a}_M(n-1)$, the residual error $e_M^f(n)$ and the error power $\alpha_M^f(n-1)$ and compute

$$\mathbf{s}_{M+1}(n) = \frac{e_M^f(n)}{\lambda\alpha_M^f(n-1)} \begin{bmatrix} 1 \\ -\mathbf{a}_M(n-1) \end{bmatrix} \quad (4)$$

(b) Using a FRLS backward predictor of order M applied to $u(n_d)$, where $n_d = n - N + M$, update the backward weight vector $\mathbf{b}_M(n_d - 1)$, the residual error $e_M^b(n_d)$ and the error power $\alpha_M^b(n_d - 1)$, and compute

$$\mathbf{t}_{M+1}(n_d) = \frac{e_M^b(n_d)}{\lambda\alpha_M^b(n_d-1)} \begin{bmatrix} -\mathbf{b}_M(n_d-1) \\ 1 \end{bmatrix} \quad (5)$$

(c) Update the dual Kalman gain $\mathbf{c}_N(n)$ and $\gamma(n)$:

$$\begin{bmatrix} \mathbf{c}_N(n) \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{c}_N(n-1) \end{bmatrix} - \begin{bmatrix} \mathbf{s}_{M+1}(n) \\ \mathbf{0}_{N-M} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{N-M} \\ \mathbf{t}_{M+1}(n_d) \end{bmatrix} \quad (6)$$

$$\gamma(n) = \gamma(n-1) + s_{M+1}^1(n)e_M^f(n) - t_{M+1}^{M+1}(n_d)e_M^b(n_d) \quad (7)$$

(c) Filtering part: Same as (2)-(3) above.

To define initial conditions for this algorithm, a soft constraint approach is described in [3]. Assuming that two distinct FAEST algorithms [5] are used in steps (a) and (b), the total complexity of FNTF is $2N + 12M$.

3. Weaver SSB subband structure

Subband adaptive filtering offers several advantages over a conventional full-band approach, including reduction of computational complexity and improved signal conditioning [2].

[1] Support for this work was provided by FCAR.