

Research Article

Drift-Compensated Adaptive Filtering for Improving Speech Intelligibility in Cases with Asynchronous Inputs

Heping Ding and David I. Havelock

Institute for Microstructural Sciences, National Research Council, 1200 Montreal Rd., Ottawa, Ontario, Canada K1A 0R6

Correspondence should be addressed to Heping Ding, heping.ding@nrc-cnrc.gc.ca

Received 4 January 2010; Revised 17 June 2010; Accepted 6 August 2010

Academic Editor: Shoji Makino

Copyright © 2010 H. Ding and D. I. Havelock. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In general, it is difficult for conventional adaptive interference cancellation schemes to improve speech intelligibility in the presence of interference whose source is obtained asynchronously with the corrupted target speech. This is because there are inevitable timing drifts between the two inputs to the system. To address this problem, a drift-compensated adaptive filtering (DCAF) scheme is proposed in this paper. It extends the conventional schemes by adopting a timing drift identification and compensation algorithm which, together with an advanced adaptive filtering algorithm, makes it possible to reduce the interference even if the magnitude of the timing drift rate is as big as one or two percent. This range is large enough to cover timing accuracy variations of most audio recording and playing devices nowadays.

1. Background

An example of the conventional adaptive interference cancellation (a.k.a. noise cancellation, or “reference canceler filter” in [1]) system is shown in Figure 1. A broadcast signal played by a TV or radio receiver in the same room as the target speech interferes with the latter and makes it less intelligible in the digitized microphone output $d(n)$. The goal is to reduce the interference $u(n)$ contained in $d(n)$ so as to improve the intelligibility of the target speech $s(n)$. To achieve this, a reference $x(n)$, being the original signal sent to the interfering loudspeaker, is filtered by an adaptive filter that automatically learns the electro-acoustic transfer function from the original to the microphone output and produces an output $y(n)$ that resembles $u(n)$. This $y(n)$ is subtracted from $d(n)$ to reduce $u(n)$ so that $s(n)$ in the output $e(n)$ is enhanced. In other words, the signal-to-interference ratio is increased.

Note that an adaptive interference cancellation system in Figure 1 or any of the others discussed in this paper is not able to reduce ambient noise uncorrelated with $x(n)$; it regards the noise as part of $s(n)$. Details about the conventional adaptive interference cancellation technology and adaptation algorithms in general can be found in [2].

With both $d(n)$ and $x(n)$ acquired synchronously—an assumption conventional schemes are based on—the system in Figure 1 may reduce the interference quite effectively.

However, in some cases, it is not easy or even possible to obtain $x(n)$ at the same time when $d(n)$ is recorded. For example, there may be restrictions so that it is only possible to place one surveillance microphone on-site and it is impossible to tap the interfering signal sent to the loudspeaker when the recording for $d(n)$ is done.

It is then suggested in Section 4.6 of [1] that one obtains the original broadcast material separately, for example, from the broadcaster, and uses it as the reference input $x(n)$. The block diagram in Figure 2 illustrates this principle. Material obtained separately may differ from the actual source of interference due to, for example, alterations or distortions during the broadcast process. As in [1], we assume in this paper that there are no such differences. In Figure 2, the broadcast material is independently played back twice—once for the interfering loudspeaker and another time when $x(n)$ is acquired. In addition, there may be more independent playback or recording operations involved during the acquisition of $d(n)$ (two more in the example of Figure 2). These operations are performed at different times and most likely by different devices.

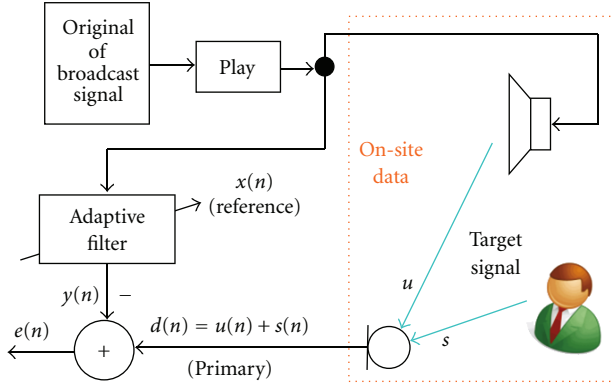


FIGURE 1: Conventional adaptive interference cancellation.

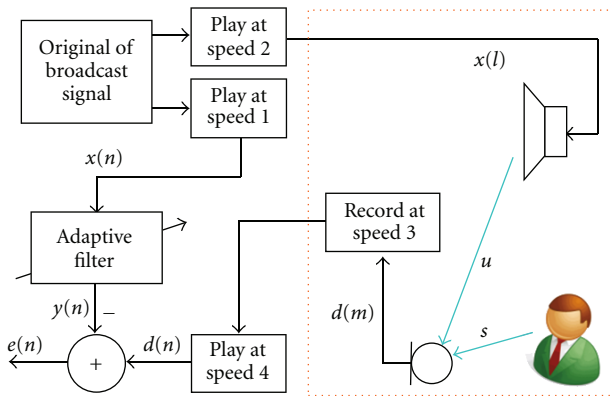


FIGURE 2: Adaptive interference cancellation with asynchronous primary and reference inputs.

It is understood that each audio recording and playback device, be it a CD player, a cassette tape recorder/player, a VHS tape recorder/player, and so forth,

- (i) records or plays at an average speed different from that of others, because of their different timing accuracies,
- (ii) has an average speed that drifts over time,
- (iii) may have irregularities in the recording/playback speed, called wow-and-flutter. This is true primarily with analog recording/playback devices.

For example, our comparison between three devices revealed that the playback speed of a consumer portable CD player is 0.066% slower than the timing provided by the sound card digitizer in a personal computer, and a higher-end DVD surround receiver plays 0.0035% slower than the sound card. The wow-and-flutter with analog devices also varies across different recorders/players and from time to time with the same recorder/player. For example, the wow-and-flutter of an analog telephone answering system is allowed to be as large as 0.3% [3]. Table 1 of [1] indicates that the speed error of an analog recording device can be as large as 3.0% and the wow-and-flutter of it 1% rms.

As a result of these factors, interference components in $d(n)$, which are supposed to be correlated with $x(n)$, are in general not synchronous with $x(n)$ in the system in Figure 2—there are varying timing drifts between them due to the differences in speeds of their respective recording and playing operations and possible timing jitters resulting from wow-and-flutter during those operations.

Note that we use l and m (instead of n) as time indices for sampled signals in the on-site data acquisition part of Figure 2. This is to emphasize the fact that they are in general played back or acquired with sampling frequencies that can differ, though slightly, from those of $\{x(n)\}$ and $\{d(n)\}$ in the adaptive filtering unit.

The asynchronous nature of the problem, together with the fact that

- (i) a misalignment—due to the timing drift—of a small fraction of a sampling interval can render a converged adaptive filter useless;
- (ii) existing adaptation algorithms usually converge much slower than these timing variations,

makes it difficult to achieve an appreciable interference reduction using just an adaptive filter in the configuration Figure 2 illustrates.

In an attempt to alleviate the adverse impact of the timing variations discussed above, it is suggested in Section 4.6 of [1] that the inputs $x(n)$ and $d(n)$ in Figure 2 be manually aligned. In practice, one may be able to compensate for a timing drift with a *constant rate* (a.k.a. linear drift) by using an interpolation/decimation means to stretch or compress the time scale of $\{x(n)\}$ or $\{d(n)\}$ according to an estimation of the drift rate, but it is a laborious process to manually estimate such a rate. Furthermore, it would take even more effort to manually look after the more general case of a timing drift with a *time-varying rate* (a.k.a. nonlinear drift). This is because $x(n)$ and $d(n)$ would first have to be partitioned into segments small enough that the drift rate during each of them can be regarded as approximately constant. Thus, manual alignment as suggested in [1] is not an effective or efficient solution to the problem. It is then necessary to find a way of automatically identifying and compensating for timing drifts regardless of whether the rates are constant or time varying.

In the application of the echo cancellation techniques to voice-over-IP networks and as software implementation on personal computers, there can be similar problems—also caused by timing variations. Examples of a software speakerphone implemented on a personal computer are in [4, 5]. The signal samples received from the far-end of a voice link are delivered to the loudspeaker(s) at a rate maybe slightly different from the rate at which the microphone signal is sampled—although these two rates are the same nominally. This situation is similar to that in Figure 2. For the acoustic echo canceller to do a decent job, it is necessary to identify the difference and compensate for it. These two algorithms focus on circumstances where the two sampling frequencies are slightly different but constant, that is, constant rate or linear drift as mentioned above.

There was extensive research in the 1980s [6, 7] on a related topic: making the echo canceller for data modems immune to certain echo-path variations. These variations were caused by a frequency shift due to slightly different carrier frequencies and by timing jitters due to coarse adjustments made by a digital phase-locked loop. It is quite effective and popular to use a phase-locked loop to estimate and compensate for the frequency shift [6], and it is possible to eliminate the adverse effect of timing jitters that happen at known time instances [7]. However, these well-developed approaches cannot be readily applied to the case in Figure 2 because the timing jitters caused by wow-and-flutter are random and unpredictable.

Thus, how to do interference cancellation in the configuration of Figure 2, with a significant and possibly time-varying timing drift between the two inputs and without any explicit information about the drift, has been an open issue. The goal of this research is to develop a scheme that is effective in this circumstance, with the expectation that it may also be applied to other applications such as those studied in [4, 5].

The rest of this paper is organized as follows: the proposed scheme is detailed in Section 2, Section 3 presents some experiment results, and Section 4 is a summary. In addition, there are three appendices that provide details of certain proofs and derivations.

2. The Proposed Scheme

In overview, the proposed drift-compensated adaptive filtering (DCAF) scheme dynamically aligns the sequence $\{d(n)\}$ with $\{x(n)\}$ by

- (i) upsampling $\{d(n)\}$ to obtain a new sequence $\{d_I(n')\}$, with a much higher time resolution;
- (ii) finding the differences (errors) between $\{d_I(n')\}$ and the adaptive filter's output;
- (iii) evaluating the errors to determine the nature of the timing drift;
- (iv) downsampling $\{d_I(n')\}$ accordingly to produce a sequence $\{d_r(n)\}$ in which the interference components are synchronous with those in $\{x(n)\}$.

The DCAF is shown in Figure 3, which is to replace the adaptive filter and the summation node in the system in Figure 2. The scheme has been briefly reported at a conference [8], and more details are provided in this paper. As illustrated, there are three major components in Figure 3: (A) timing drift estimation and compensation, which is the essence of the proposed scheme and looks after the time alignment between the two inputs; (B) Ratchet fast affine projection (FAP) adaptive filter, for fast convergence and low complexity; and (C) peak position adjustment, which is indispensable for such a time-drifting application of adaptive filtering. These three components will be discussed separately below.

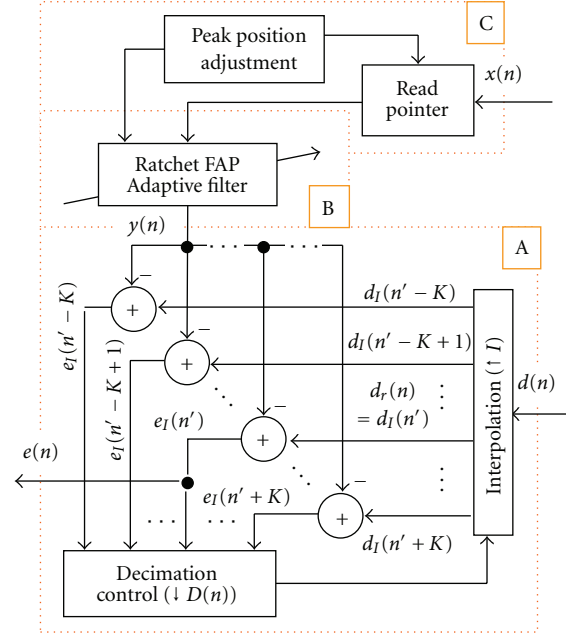


FIGURE 3: Proposed DCAF scheme.

In this paper, we only discuss the time-domain approach for ease of understanding the concepts. In practice, the DCAF could also be implemented in the frequency domain for improved efficiency.

2.1. Timing Drift Estimation and Compensation. The term “timing drift” will henceforth refer to the aggregated net effect of timing variations resulting from all playback and recording operations involved, such as those in Figure 2. In the DCAF scheme, the timing drift is dynamically estimated by evaluating certain time averages and then compensated for by properly resampling the primary input sequence $\{d(n)\}$ to form a new sequence $\{d_r(n)\}$ in which the interference components are synchronous with the reference input sequence $\{x(n)\}$. In other words, the sampling frequency for $\{d(n)\}$ is dynamically adjusted so that the resultant $\{d_r(n)\}$ has the same sampling frequency as that of $\{x(n)\}$ —as if $\{d_r(n)\}$ and $\{x(n)\}$ were acquired synchronously. That being done, the adaptive filter is able to make a reliable estimate of the interference in $\{d_r(n)\}$. We now look at how the resampling is implemented, how the timing drift is estimated, and how the resampling is controlled to compensate for the timing drift.

To resample $\{d(n)\}$, it is first upsampled by a factor I ($I = 100$ in this paper), resulting in an interpolated sequence $\{d_I(n')\}$:

$$\begin{aligned} \dots, d_I(nI - 1), d_I(nI) \approx d(n), d_I(nI + 1), \dots, \\ d_I((n + 1)I - 1), d_I((n + 1)I) \approx d(n + 1), \\ d_I((n + 1)I + 1), \dots \end{aligned} \quad (1)$$

whose sampling frequency F_{SI} is I times that of $\{d(n)\}$. This is illustrated in Figure 4.

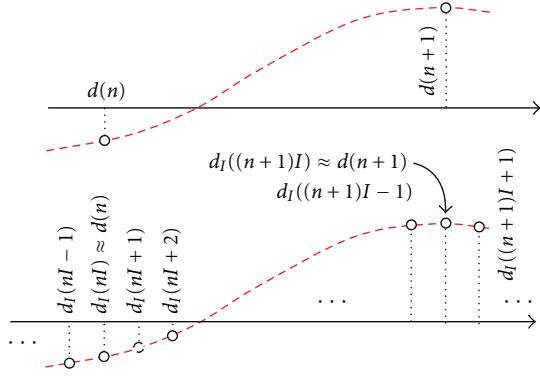


FIGURE 4: Upsampling $d(n)$ to get $d_I(n')$.

The upsampling is performed by first padding $I - 1$ zeros between each pair of adjacent samples in $\{d(n)\}$ then passing the resultant sequence through a low-pass filter. In the case used in our experiments, $I = 100$, and the FIR low-pass filter has 10208 coefficients, which are symmetric so that the filter has a frequency-independent group delay of $(10208 - 1)/2 = 5103.5$ interpolated samples. The passband ripple and stopband attenuation are 0.5 dB and 50 dB, respectively. The passband and stopband edges are located at $0.0048125 F_{SI}$ and $0.005 F_{SI}$, respectively. Details about upsampling techniques can be found in a text book on digital signal processing, for example, [9].

Then, $\{d_I(n')\}$ is decimated by a time-varying factor $D(n) \approx I$ to arrive at the resampled sequence $\{d_r(n)\}$, whose sampling frequency approximately equals that of $\{d(n)\}$. This is achieved by

$$d_r(n) = d_I(n'), \quad (2)$$

where

$$n' \equiv (n + \Delta)I + [\text{offset}(n)]. \quad (3)$$

In (3), Δ is an integer, $[\cdot]$ denotes the rounding operation, and $0 \leq \text{offset}(n) < I$. Thus, $d_r(n)$ leads $d(n)$ by $\Delta + [\text{offset}(n)]/I$ original (not upsampled) samples.

If $\text{offset}(n)$ has a constant value, then $D(n) \equiv I$; that is, $\{d_r(n)\}$ and $\{d(n)\}$ have the same sampling frequency but may have a constant offset in time. However, a time-varying $\text{offset}(n)$ may result in $D(n)$ deviating from I .

The key to timing drift compensation is to dynamically adjust $D(n)$ by modifying $\text{offset}(n)$ in (3) so that the interference components in $\{d_r(n)\}$ stay synchronous with $\{x(n)\}$. To do so, we update $\text{offset}(n)$ adaptively using

$$\text{offset}(n+1) = \text{offset}(n) + \text{offset_inc}(n), \quad (4)$$

where the updating term $\text{offset_inc}(n)$ stands for “offset increment.” When the right-hand side of (4) goes beyond the

range $[0, I - 1]$, wraparound is performed as follows

If $\text{offset}(n+1) \geq I$, then

$$\text{offset}(n+1) = \text{offset}(n+1) - I, \quad \Delta = \Delta + 1. \quad (5)$$

Else if $\text{offset}(n+1) < 0$, then

$$\text{offset}(n+1) = \text{offset}(n+1) + I, \quad \Delta = \Delta - 1$$

so that $\text{offset}(n+1)$ remains in the range $[0, I - 1]$.

Based on (2)–(4), the decimation factor is

$$D(n) \equiv \frac{\partial n'}{\partial n} = I + \frac{\partial[\text{offset}(n)]}{\partial n} = I + \text{offset_inc}(n) + \delta, \quad (6)$$

where δ is a zero-mean noise resulting from rounding; therefore, its rms value is $1/(2\sqrt{3})$. In a steady state, for example, the timing drift rate is constant (the case considered in [4, 5]), and $D(n)$ is expected to wobble around a constant defined by $\langle D(n) \rangle = I + \langle \text{offset_inc}(n) \rangle$, where $\langle \cdot \rangle$ is the time-averaging operator. It follows that, in that case, the ratio between the sampling frequencies of the original and the resampled sequences is

$$\frac{\langle D(n) \rangle}{I} = 1 + \frac{\langle \text{offset_inc}(n) \rangle}{I}. \quad (7)$$

The remaining issue is to estimate the timing drift so as to control $\text{offset_inc}(n)$. We begin with a $(2K + 1)$ -element ($K < I/2$) subsequence:

$$\{d_I(n' + k), \forall k \in [-K, K]\} \quad (8)$$

of (1). In (8), K typically equals 15 in our experiments, and wraparound adjustments as per (5) are made if any $\text{offset}(n) + k$ becomes out of $[0, I - 1]$. Note that the element in the middle of (8) is (2).

As illustrated in Figure 3, the adaptive filter’s output $y(n)$ is subtracted from (8) to produce $2K + 1$ error values

$$e_I(n' + k) = d_I(n' + k) - y(n), \quad \forall k \in [-K, K], \quad (9)$$

with the main error value in the middle at $k = 0$. This enables us to examine the output error with an I -times finer time resolution—to facilitate timing drift estimation.

Let us consider the expectations

$$E[e_I^2(n' + k)], \quad \forall k \in [-K, K]. \quad (10)$$

It is henceforth assumed that the adaptive filter has mostly converged and there exists a unique $k_{\text{opt}} \in [-K, K]$ so that

$$E[e_I^2(n' + k_{\text{opt}})] < E[e_I^2(n' + k)], \quad \forall k \in [-K, K]. \quad (11)$$

It is proven in Appendix A that elements in (10) form a convex and approximately quadratic function of k if $|k - k_{\text{opt}}| < I/2$ and the target signal $s(n)$ plus the ambient noise are uncorrelated with $x(n)$.

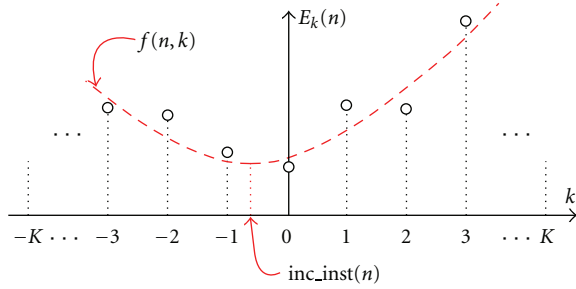


FIGURE 5: Least-squares curve fitting.

We then need to control `offset_inc(n)` in (4) for consecutive sampling intervals in order for the main (middle) error $e_I(n')$ to remain at the minimum in (11); that is, $k_{\text{opt}} = 0$. Thus, it is necessary to monitor (10) and keep track of the actual position of its minimum. Since it is impossible to find ensemble means in practice, (10) has to be approximated, for example, by time averages. What we adopt is (12), with first-order smoothing over time:

$$E_k(n) = \beta E_k(n-1) + (1-\beta) e_I^2(n'+k), \quad \forall k \in [-K, K], \quad (12)$$

where $\beta \in (0, 1)$ is close to 1. Note that the relation between the time indices n and n' in (12) is defined by (3). Next, a parabola $f(n, k)$ that fits the elements in (12) in the least-squares sense is found. If $f(n, k)$ is convex as expected, then a finite minimum `inc_inst(n)` of it exists, as illustrated in Figure 5. It is shown in Appendix B that $f(n, k)$ is convex if

$$a'(n) \equiv 3 \sum_{k=-K}^K k^2 E_k(n) - K(K+1) \sum_{k=-K}^K E_k(n) > 0, \quad (13)$$

and, in that case,

$$\text{inc_inst}(n) = \frac{4K^2 + 4K - 3}{-10a'(n)} \sum_{k=-K}^K k E_k(n). \quad (14)$$

This is a candidate for `offset_inc(n)`.

Due to the presence of the target signal $s(n)$, the ambient noise, and uncancelable interference,

- (i) equation (14) may be too noisy to be used as `offset_inc(n)` in (4);
- (ii) it is possible for $f(n, k)$ to be nonconvex—indicated by (13) as being not satisfied. If so, (14) is not meaningful.

Thus, the `offset_inc(n)` is found by using a smoothing operation over many sampling intervals:

$$\begin{aligned} \text{offset_inc}(n) &= \text{offset_inc}(n-1) \\ &+ \begin{cases} \mu \cdot \text{inc_inst}(n) & \text{if } a'(n) > 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (15)$$

where μ is a small positive step size.

Finally, the interference-reduced system output is the main error in (9); that is,

$$e(n) \equiv e_I(n') = d_r(n) - y(n). \quad (16)$$

We now address the issue of selecting the interpolation factor I . As seen, the resolution of the timing drift compensation is $1/I$ of a sampling interval. For the sake of reducing implementation complexity, a small value for I is beneficial. It is then necessary to find a smallest I without sacrificing the perceptible cancellation performance. Through some manipulations, Appendix C gives the following guideline:

$$I > \pi \cdot 10^{\text{TR}/20}, \quad (17)$$

where TR is the wanted ratio (in dB) of the level of $d(n)$ to the level of tolerable adjustment errors; that is, the errors should be TR dB lower in level than the primary input. Experiments suggest that TR = 30 dB, which results in $I = 100$, gives an adequate tradeoff between performance and complexity.

Note that, although $2K+1$ errors are calculated in (9), the added complexity is quite small since there is only one adaptive filter. Another remark is that the upsampling of $\{d(n)\}$ by a seemingly large factor of $I = 100$ is mainly conceptual. In reality, only $2K+1$ interpolated values in (8)—as opposed to all those in (1)—need to be calculated and, for each of them, 99% (for $I = 100$) of the input samples to the 10208-coefficient FIR interpolation filter are zeros. Thus, the polyphase filtering technique [9] is adopted so that the computation load is minimized.

2.2. Ratchet FAP. Although any adaptive filter could potentially be used in Figure 3, one adopting the Ratchet FAP algorithm [10] is chosen. This is because (a) a FAP can converge an order of magnitude faster than the most commonly used NLMS and is only marginally more complex; and (b) the Ratchet FAP is superior to other FAP algorithms in terms of performance and stability. In addition to adaptive interference cancellation, Ratchet FAP can also find applications in echo cancellation, source separation [11], hearing aids, and other areas in communications and medical signal processing.

The Ratchet FAP used in this application incorporates an algorithm that dynamically optimizes the regularization factor so that it is just large enough to assure stability of the implicit matrix inversion process associated with the FAP. See [12] for further information.

2.3. Peak Position Adjustment. An important issue with such a time-drifting application of adaptive filtering is that the coefficients of the adaptive filter may drift over time, even after convergence. Corresponding approximately to the filter's group delay, the *main part* of the coefficients that needs to be considered is typically a small, contiguous set of coefficients with large magnitudes. If this part moves close to the beginning or end of the range spanned by the adaptive filter, the interference reduction performance may significantly degrade.

To circumvent this, the position of the main part of the coefficients is constantly monitored and adjustments are performed when necessary. This position is estimated by

$$\text{pos}_m(n, q) = \frac{\sum_{k=1}^{L-1} k |w_k(n)|^q}{\sum_{k=0}^{L-1} |w_k(n)|^q} \quad (18)$$

in a manner similar to how “center of gravity” is estimated. In (18), the subscript m stands for “main,” and $\{w_0(n), w_1(n), \dots, w_{L-1}(n)\}$ are the L coefficients of the Ratchet FAP adaptive filter in Figure 3. Equation (18) with the parameter $q = 1$ gives the position of the center of magnitudes (center of mass), with $q = 2$ gives the center of energy (moment of inertia) or the filter’s group delay, and with $q = \infty$ gives the index of the coefficient with the largest magnitude. In our experiments, $q = 4$ is used in order to take into account both the group delay and large peaks.

Next, (18) is compared against a target range of values that can be determined heuristically. If the deviation is significant enough, then realignment adjustments, with a step of one sample every preset number of sampling intervals, are made until the deviation lies within the target range. The realignment adjustments require changes to

- (i) the read pointer for $x(n)$ (Figure 3);
- (ii) the coefficients of the adaptive filter—they are shifted one sample to the left or right (depending on the need) with a zero appended to the opposite end;
- (iii) the autocorrelation matrix estimate of the Ratchet FAP adaptive filter—the sums therein need also to be shifted and properly appended accordingly.

Further incidental implementation details are needed but these are omitted here for brevity.

A remark about the read-pointer adjustment mentioned above is that, in a real-time implementation, such adjustments may result in serious consequences as over- or underflow of the input buffers can occur. This problem is common in telecommunications (see Section 1), and there are techniques to circumvent it. However, this topic is beyond the scope of this paper; our purpose is to propose an algorithm’s framework, and all processings presented in Section 3 have been done offline so that the over- or underflow issue is avoided.

2.4. About Adaptation Control. It is normally necessary for an adaptive system such as the DCAF to have an adaptation control to prevent the adaptive systems from potentially diverging when the target signal s is active. This could be done by nullifying the two step sizes, for example, μ in (15) and that for the Ratchet FAP. The detection of this condition is called “double-talk detection” in literature on echo cancellation.

Contrary to this, no adaptation control is implemented in the current DCAF scheme because, in this application (see Section 1), the interference and target can be active simultaneously most of the time. This leaves very little “single-talk” (no target) time in which the adaptation

TABLE 1: DCAF’s performance without and with timing drift compensation—simulated conditions.

Test case	Nature of timing drift rate during the 120 s test case period	Achieved interference reduction (dB) with timing drift compensation being	
		disabled	enabled
1	0 \nearrow 1% linearly	1.2	7.5
2	0 \searrow -1% linearly	0.4	11.3
3	0 \nearrow 1% \searrow 0 linearly	1.6	9.4
4	0 \searrow -1% \nearrow 0 linearly	0.3	8.6
5	1/60 Hz cosine with peaks $\pm 0.08\%$	2.6	9.4

systems could adapt quickly and reliably. Indeed, the system the DCAF tries to approximate is expected to change only slowly, and so the adaptation is allowed to take place full-time (i.e., even during double talk) but with very small step sizes. The resultant DCAF scheme is a compromise between convergence speed and immunity to the target signal. It could be a future research topic to find a way of optimally controlling the step sizes in conjunction with double-talk detection.

3. Experiments

The proposed DCAF scheme has been evaluated with real-room signals combined under simulated conditions. The real-room signals use recording and playback devices having different timing accuracies. The sampling frequencies used are (nominally) 8, 16, 44.1, and 48 kHz.

Subjective evaluation to characterize the intelligibility improvement has been performed. Its process and results are reported in Section 3.3.

3.1. Simulated Conditions. Test cases are prepared using recorded radio broadcast signals filtered with 740 ms long room impulse responses which were measured in a large meeting room. The timing drifts are created by properly controlled resampling and delaying of the primary or reference input.

Table 1 lists several test cases, all with a 16 kHz sampling frequency, a 120-second duration, and a signal-to-interference ratio in $\{d(n)\}$, before processing, of -1.4 dB. In the DCAF scheme, the Ratchet FAP adaptive filter has $L = 2000$ coefficients (125 ms) and an affine projection order $N = 5$. The normalized step size α of the adaptive filter starts with a relatively large value of 0.050–0.100 and diminishes to 0.005–0.010 after initial convergence. In the drift compensation part, the interpolation factor is $I = 100$, the parameter $K = 15$, and the step size μ in (15) is either equal to 0 or in the approximate range of $5 \times 10^{-6} \sim 10^{-5}$. When $\mu = 0$, the drift compensation part (Section 2.1) is disabled so that the DCAF falls back to a conventional adaptive interference cancellation scheme.

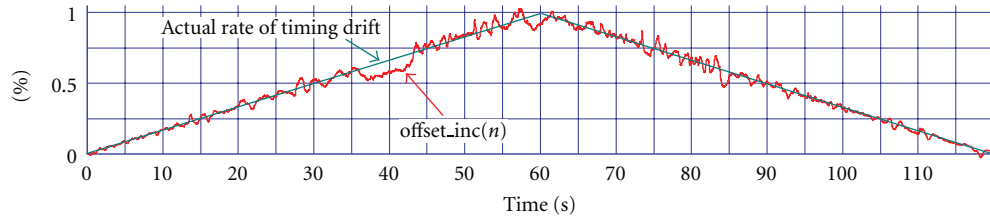


FIGURE 6: Actual and estimated rates of timing drift for Test Case 3.

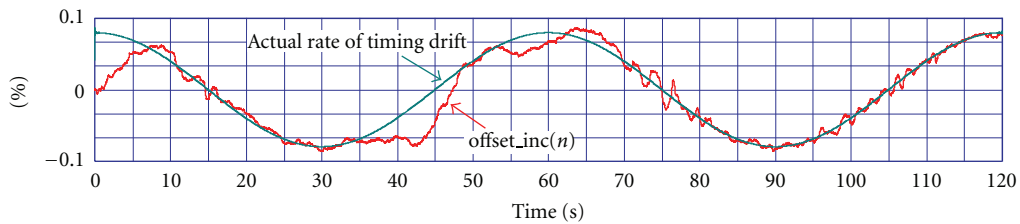


FIGURE 7: Actual and estimated rates of timing drift for Test Case 5.

Note that, in order to estimate the amount of interference reduction accurately, the energy (sum of squares of all samples over the entire test case period) of the target signal (which is known since simulated conditions are dealt with) is subtracted from energies of $\{d_r(n)\}$ and $\{e(n)\}$ before figures in Table 1 are calculated.

Table 1 indicates that the DCAF scheme can reduce the interference by 7–11 dB. When the drift compensation part is disabled, the DCAF falls back to a conventional algorithm. In that case, it is not capable of handling these timing drifts. Consequently, little interference reduction is observed, as shown in Table 1.

Consider Test Case 3 in Table 1 as an example. The rate of the timing drift between the two inputs goes linearly from 0 to 1% in 60 seconds and back to 0, again linearly, in the next 60 seconds. Figure 6 shows that the DCAF has correctly estimated that rate.

In Test Case 5, another example, the rate of the timing drift between the two inputs varies according to a sinusoidal pattern. It can be seen in Figure 7 that it takes some time for the DCAF to initially catch up to the timing drift. Once the initial alignment has been achieved, the algorithm stays in synchronization.

It is clearly seen in Figures 6 and 7 that the $offset_inc(n)$ is still quite noisy despite the smoothing operations (12) and (15). This phenomenon has also been observed in other test cases in Table 1. This is believed to be attributed to the presence of the strong target signal plus ambient noise (only 1.4 dB below the interference) and uncancelable interference—as discussed in Section 2.4. This will be verified by the next test case in Section 3.2.

3.2. Real Room with Real Recording and Playback Devices. With the primary input recorded in real rooms by real recording and playback devices having different speeds, these tests aim at verifying the performance of the DCAF in real life.

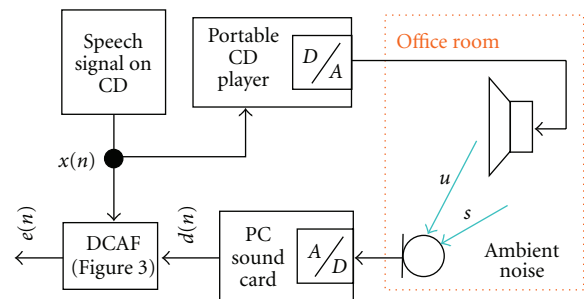


FIGURE 8: A room recording setup.

Figure 8 illustrates the recording setup in an ordinary office room. The portable CD player plays the digitally stored interfering speech $x(n)$ at a slightly lower sampling rate than that of the PC sound card used to digitize the primary input to get $d(n)$. In this test scenario, the target signal s is the steady ambient noise, resulting mostly from equipment and ventilation fans in the room. It has a level 19 dB below that of the interference x introduced by the loudspeaker. The primary input $d(n)$ is sampled at 8 kHz and has a duration of 900 seconds. In the DCAF, the Ratchet FAP adaptive filter has $L = 1000$ coefficients (125 ms) and a step size $\alpha = 0.05$ throughout the entire period. It is observed that the interference reduction is only 2.1 dB if $\mu = 0$ (drift compensation disabled) and reaches 19.3 dB if $\mu = 5 \times 10^{-6}$. Figure 9 shows that after a few seconds of initial learning the DCAF estimates a timing drift rate of around 0.066%, and this value rises slightly to around 0.07% towards the end of the run. This rising is thought to correspond to the variation of the actual timing drift rate over the 900-second period. In this test case, the target signal plus the ambient noise and the uncancelable interference are much lower in level than was the case in Section 3.1. This explains why the estimate for $offset_inc(n)$ is much less noisy.

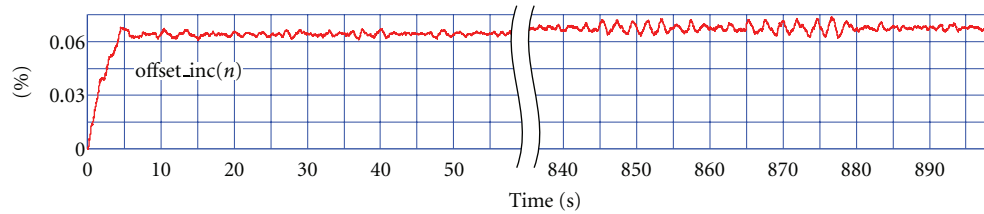


FIGURE 9: Estimated rate of timing drift for room recording with ambient noise but no target signal.

With other real-life signals, recorded in rooms and by devices different from those used for Figure 8, the interference reduction is consistent with the cases with simulated conditions (Section 3.1) when the magnitude of the rate of the timing drift is not very large, for example, no more than 0.5%.

When an analog cassette audio recorder/player is used, the observed magnitude of the varying timing drift rate can be as large as 3%. It has been observed (but not reported in detail here) that, although the DCAF still converges and tracks the drift, the interference-reduction performance degrades when the timing drift rate reaches such a large magnitude. For example, the interference-reduction can be only around 1 or 2 dB and is barely perceivable by human ears. It is believed that the relatively severe wow-and-flutter of the particular analog device used, not just the large magnitude of the timing drift rate, may likely have contributed to the performance degradation. Fortunately, wow-and-flutter is virtually nonexistent with modern digital devices.

3.3. Subjective Evaluation. To assess the performance of the proposed DCAF scheme in terms of improved intelligibility, subjective tests were conducted with 25 individuals. The intelligibility of test signals is compared for three processing conditions: (a) no processing, (b) processing with the DCAF, and (c) processing conducted by an acoustic forensic expert using conventional methodologies.

The test signals consist of target male-spoken English sentences (the IEEE “Harvard sentences” [13]) with interfering speech babble. The target and interfering signals are processed through room impulse responses from different locations within the same room and then mixed to a specified signal-to-interference ratio (SIR). A time-varying timing drift is applied to the mixed signals using two drift patterns: a sinusoidal variation with a period of 60s and peak change in sampling rate of 0.04% and a pseudorandom variation with peaks of about 0.025%. These timing drifts are imperceptible to normal listening but have a significant impact on conventional interference cancellation.

The leading and trailing portions of the processed test signals are discarded to ensure algorithm convergence and avoid any possible end effects. To examine the variety of test conditions, each subject is presented with 100 randomized test sentences. Each test sentence is padded with interference to a fixed duration of 4.5 s. After listening to each sentence, the subject repeats back the words that were understood and the fraction of words correct is recorded.

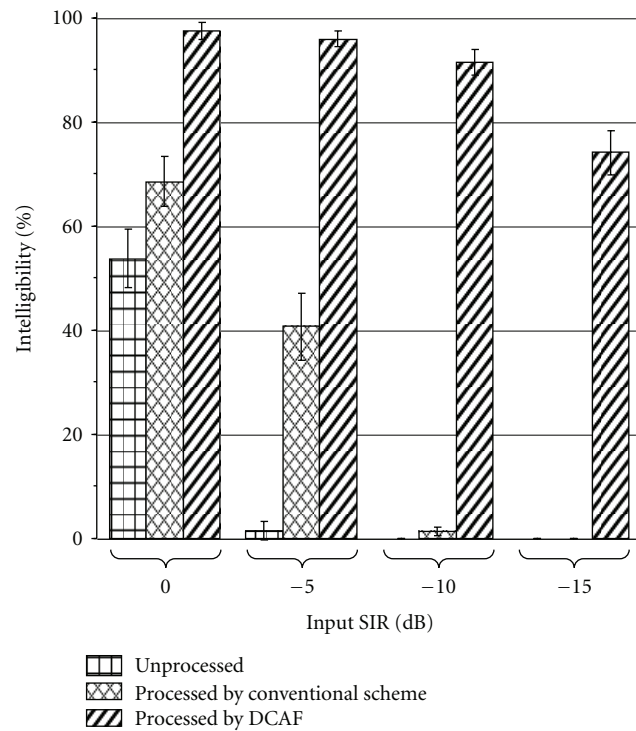


FIGURE 10: Intelligibility with three processing conditions.

The resulting intelligibility is shown in Figure 10 as a percentage of words correctly understood, for the selected SIR values and the three processing conditions. Error bars indicate the standard deviation of observed data. At all tested SIR, the proposed DCAF scheme provided very good intelligibility even though the conventional processing provided little or no intelligibility improvement at lower SIR.

3.4. Some Discussions. The DCAF algorithm can, in principle, accommodate any timing variation between the reference and primary inputs as long as it is relatively slow. Therefore, there should be a limit on the rate of acceleration or deceleration of the timing drift (i.e., rate at which the timing drift rate varies) that the DCAF can track. Although there are no comprehensive characterization data available at this time, observations suggest that the DCAF can achieve noticeable interference reduction for acceleration rates as large as $\pm 1\%$ per 60 seconds at a 16 kHz sampling rate, as seen in Test Cases 3 and 4 in Table 1. In other words, the timing drift rate changes by 1% over a period of

60 × 16000 samples. A way of expressing the magnitude of this acceleration of the timing drift (in “units” of “offset in samples”/sample²) is

$$\frac{1\%}{60 \times 16000} \approx 1.04 \times 10^{-8} \text{ sample}^{-1}. \quad (19)$$

Increasing the step size μ in (15) to a value beyond that used in our experiments, which is 5×10^{-6} , may improve the above tracking performance index, but at the expense of reduced noise immunity of the DCAF.

4. Summary

By adopting a unique estimation and compensation mechanism, a drift-compensated adaptive filtering (DCAF) scheme is proposed. The scheme makes it possible for an adaptive interference canceller to survive time-varying timing drifts between the two inputs to a degree large enough to accommodate timing accuracy variations of most audio recording and playing devices nowadays. On the contrary, conventional schemes typically fail completely under conditions of even small timing drifts. The DCAF scheme is suitable for applications in which the reference and primary inputs may be asynchronous with each other. Example applications include certain surveillance scenarios, network echo cancellation for voice-over-IP networks, and software acoustic echo cancellation implemented on personal computers.

Appendices

A. Convexity and Quadraticity

We now prove that, as long as the system in “A” of Figure 3 is slowing time-varying, elements in (10) form a convex and approximately quadratic function of k if (a) the adaptive filter has mostly converged, (b) the target signal $s(n)$ plus the ambient noise are uncorrelated with $x(n)$, and (c)

$$|k - k_{\text{opt}}| < 0.5I. \quad (A.1)$$

For convenience, we define $\Delta k \equiv k - k_{\text{opt}}$.

Equation (11) indicates that the interference components in $d_I(n' + k_{\text{opt}})$ are well aligned with $y(n)$. As a result, $d_I(n' + k_{\text{opt}})$ can be expressed as

$$d_r(n) \equiv d_I(n' + k_{\text{opt}}) = y(n) + v(n), \quad (A.2)$$

where the noise $v(n)$ is uncorrelated to $y(n)$ and consists of the target signal $s(n)$, the ambient noise, and uncancelable interference.

The discrete-time Fourier transforms of $y(n)$ and $v(n)$ in (A.2) are

$$Y(\omega) = \sum_{n=-\infty}^{\infty} y(n)e^{-j\omega n}, \quad V(\omega) = \sum_{n=-\infty}^{\infty} v(n)e^{-j\omega n}, \quad (A.3)$$

and $y(n)$ and $v(n)$ can be expressed as inverse transforms

$$y(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} Y(\omega)e^{j\omega n} d\omega, \quad v(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} V(\omega)e^{j\omega n} d\omega. \quad (A.4)$$

It follows that (8), being interpolated from $d(n)$, can be written as

$$\begin{aligned} d_I(n' + k) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} Y(\omega)e^{j\omega(n+\Delta k/I)} d\omega \\ &+ \frac{1}{2\pi} \int_{-\pi}^{\pi} V(\omega)e^{j\omega(n+\Delta k/I)} d\omega, \end{aligned} \quad (A.5)$$

$$\forall k \in [-K, K].$$

Therefore, (9) can be expressed as

$$\begin{aligned} e_I(n' + k) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} Y(\omega)(e^{j\omega\Delta k/I} - 1)e^{j\omega n} d\omega \\ &+ \frac{1}{2\pi} \int_{-\pi}^{\pi} V(\omega)e^{j\omega(n+\Delta k/I)} d\omega, \end{aligned} \quad (A.6)$$

$$\forall k \in [-K, K].$$

Given $y(n)$ and $v(n)$ being uncorrelated, (10) becomes

$$\begin{aligned} E[e_I^2(n' + k)] &= \frac{1}{4\pi^2} \iint_{-\pi}^{\pi} E[Y(\omega)Y^*(\omega')]e^{j(\omega-\omega')n} \\ &\cdot (e^{j(\omega-\omega')\Delta k/I} + 1 - e^{j\omega\Delta k/I} - e^{-j\omega'\Delta k/I}) d\omega d\omega' \quad (A.7) \\ &+ \frac{1}{4\pi^2} \iint_{-\pi}^{\pi} E[V(\omega)V^*(\omega')]e^{j(\omega-\omega')(n+\Delta k/I)} d\omega d\omega', \\ &\forall k \in [-K, K], \end{aligned}$$

where the superscript (*) denotes complex conjugate.

To simplify (A.7), we use

$$\begin{aligned} E[Y(\omega)Y^*(\omega')] &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} E[y(n)y(m)]e^{-j\omega n}e^{j\omega' m} \\ &= \sum_{m=-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} R_y(n-m)e^{-j\omega(n-m)} \right] \\ &\quad \times e^{-j(\omega-\omega')m}, \end{aligned} \quad (A.8)$$

where

$$R_y(l) \equiv E[y(n)y(n+l)] \quad (A.9)$$

is the autocorrelation function of $y(n)$. By letting $l \equiv n - m$, (A.8) becomes

$$\begin{aligned} E[Y(\omega)Y^*(\omega')] &= \sum_{m=-\infty}^{\infty} \left[\sum_{l=-\infty}^{\infty} R_y(l)e^{-j\omega l} \right] e^{-j(\omega-\omega')m} \\ &= S_y(\omega) \sum_{m=-\infty}^{\infty} e^{-j(\omega-\omega')m} = 2\pi S_y(\omega)\delta_{\omega-\omega'}, \end{aligned} \quad (A.10)$$

where δ_{ω} is the Dirac delta function of ω and

$$S_y(\omega) \equiv \sum_{l=-\infty}^{\infty} R_y(l)e^{-j\omega l}. \quad (A.11)$$

Similarly, for the noise we have

$$E[V(\omega)V^*(\omega')] = 2\pi S_v(\omega)\delta_{\omega-\omega'}, \quad (\text{A.12})$$

where

$$S_v(\omega) \equiv \sum_{l=-\infty}^{\infty} R_v(l)e^{-j\omega l}, \quad R_v(l) \equiv E[v(n)v(n+l)]. \quad (\text{A.13})$$

Substituting (A.10) and (A.12) into (A.7) results in

$$E[e_I^2(n'+k)] = \frac{2}{\pi} \int_{-\pi}^{\pi} S_y(\omega) \sin^2\left(\frac{\Delta k}{2I}\omega\right) d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} S_v(\omega) d\omega, \quad \forall k \in [-K, K]. \quad (\text{A.14})$$

Given (A.1) and $|\omega| \leq \pi$ in (A.14), the argument of the sine function here is quite small in magnitude; therefore,

$$\sin\left(\frac{\Delta k}{2I}\omega\right) \approx \frac{\Delta k}{2I}\omega, \quad (\text{A.15})$$

and (A.14) can be written as

$$E[e_I^2(n'+k)] \approx \frac{(k - k_{\text{opt}})^2}{2\pi I^2} \int_{-\pi}^{\pi} S_y(\omega) \omega^2 d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} S_v(\omega) d\omega, \quad \forall k \in [-K, K]. \quad (\text{A.16})$$

While (11) only requires that there be a minimum at $k = k_{\text{opt}}$, (A.16) further shows that elements in (10) form a convex and approximately quadratic function of k .

B. Least Squares Curve Fitting

Here, we prove the validity of (13) and (14).

The parabolic curve $f(n, k)$ illustrated in Figure 5 can be defined by parameters $\{a(n), b(n), c(n)\}$ as in

$$f(n, k) = a(n)k^2 + b(n)k + c(n), \quad \forall k \in [-K, K]. \quad (\text{B.1})$$

To find the parameters that make (B.1) approximate the $2K + 1$ estimates in (12) in a least-squares sense, we minimize the nonnegative cost function

$$C(n) = \sum_{k=-K}^K [f(n, k) - E_k(n)]^2 \quad (\text{B.2})$$

by letting its partial derivatives with respect to the three parameters $\{a(n), b(n), c(n)\}$ be zeros. This leads to a system of linear equations

$$\begin{bmatrix} S_4 & S_3 & S_2 \\ S_3 & S_2 & S_1 \\ S_2 & S_1 & 2K+1 \end{bmatrix} \begin{bmatrix} a(n) \\ b(n) \\ c(n) \end{bmatrix} = \begin{bmatrix} T_2(n) \\ T_1(n) \\ T_0(n) \end{bmatrix}, \quad (\text{B.3})$$

where

$$S_m \equiv \sum_{k=-K}^K k^m, \quad T_m(n) \equiv \sum_{k=-K}^K k^m E_k(n). \quad (\text{B.4})$$

The antisymmetry property makes $S_m = 0$, for all m odd; therefore, (B.3) simplifies to

$$b(n) = \frac{T_1(n)}{S_2}, \quad (\text{B.5})$$

$$\begin{bmatrix} S_4 & S_2 \\ S_2 & 2K+1 \end{bmatrix} \begin{bmatrix} a(n) \\ c(n) \end{bmatrix} = \begin{bmatrix} T_2(n) \\ T_0(n) \end{bmatrix}.$$

Given that

$$S_2 = K(K+1)(2K+1)/3, \quad (\text{B.6})$$

$$S_4 = K(K+1)(2K+1)(3K^2+3K-1)/15,$$

one can solve (B.5) to get

$$a(n) = \frac{15}{K(K+1)(2K+1)(4K^2+4K-3)} a'(n), \quad (\text{B.7})$$

where

$$a'(n) \equiv 3T_2(n) - K(K+1)T_0(n). \quad (\text{B.8})$$

The fact that (B.7) and (B.8) (which is equivalent to (13)) are positive indicates that (B.1) is convex. If so, a finite minimum of (B.1) exists and is at

$$\text{inc_inst}(n) \equiv \frac{-b(n)}{2a(n)} = \frac{4K^2+4K-3}{-10} \cdot \frac{T_1(n)}{a'(n)}, \quad (\text{B.9})$$

which is (14).

C. Choosing Interpolation Factor

We now study how to choose the interpolation factor I based on how adjustment errors resulting from it degrade the noise performance of the DCAF scheme.

The resolution of the timing drift compensation is $1/I$ of a sampling interval, so we must choose I to be large enough that k fluctuating by ± 1 in the vicinity of $k = k_{\text{opt}}$ does not lead to a perceptibly significant performance degradation. This is expressed as

$$E\left\{ \left[e_I(n' + k_{\text{opt}} \pm 1) - e_I(n' + k_{\text{opt}}) \right]^2 \right\} < \sigma_T^2, \quad (\text{C.1})$$

where σ_T^2 is the tolerable power of the adjustment errors. For example, if σ_T^2 is below a just-noticeable threshold, (C.1) assures that a ± 1 error in k around k_{opt} is not audible.

Given (9), (C.1) is actually

$$E\{(\Delta d)^2\} < \sigma_T^2, \quad (\text{C.2})$$

where $\Delta d \equiv d_I(n' + k_{\text{opt}} \pm 1) - d_I(n' + k_{\text{opt}})$. Using the Fourier transform pair

$$D_r(\omega) = \sum_{n=-\infty}^{\infty} d_r(n) e^{-j\omega n} \equiv \sum_{n=-\infty}^{\infty} d_I(n' + k_{\text{opt}}) e^{-j\omega n}, \quad (\text{C.3})$$

$$d_r(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} D_r(\omega) e^{j\omega n} d\omega$$

and following the same rationale as in (A.5), we get

$$\begin{aligned}\Delta d &= \frac{1}{2\pi} \int_{-\pi}^{\pi} D_r(\omega) e^{j\omega n} (e^{\pm j\omega/I} - 1) d\omega \\ &= \pm \frac{j}{\pi} \int_{-\pi}^{\pi} D_r(\omega) e^{j\omega(n \pm 0.5/I)} \sin \frac{\omega}{2I} d\omega\end{aligned}\quad (\text{C.4})$$

so that

$$\begin{aligned}E\{(\Delta d)^2\} &= \frac{1}{\pi^2} \iint_{-\pi}^{\pi} E[D_r(\omega) D_r^*(\omega')] \\ &\quad \times \sin \frac{\omega}{2I} \sin \frac{\omega'}{2I} e^{j(\omega - \omega')(n \pm 0.5/I)} d\omega d\omega'.\end{aligned}\quad (\text{C.5})$$

Similar to (A.9) through (A.11), we can write

$$E[D_r(\omega) D_r^*(\omega')] = 2\pi S_d(\omega) \delta_{\omega - \omega'}, \quad (\text{C.6})$$

where

$$S_d(\omega) \equiv \sum_{l=-\infty}^{\infty} E[d_r(n) d_r(n+l)] e^{-j\omega l}. \quad (\text{C.7})$$

Substituting (C.6) into (C.5) results in

$$E\{(\Delta d)^2\} = \frac{2}{\pi} \int_{-\pi}^{\pi} S_d(\omega) \sin^2 \frac{\omega}{2I} d\omega. \quad (\text{C.8})$$

Considering $|\omega/(2I)| < \pi/4$ for any reasonable choice of I ($I > 2$) and using (C.7), (C.8) can be written as

$$\begin{aligned}E\{(\Delta d)^2\} &\approx \frac{1}{2\pi I^2} \int_{-\pi}^{\pi} S_d(\omega) \omega^2 d\omega \\ &= \frac{1}{2\pi I^2} \sum_{l=-\infty}^{\infty} E[d_r(n) d_r(n+l)] \int_{-\pi}^{\pi} \omega^2 e^{-j\omega l} d\omega.\end{aligned}\quad (\text{C.9})$$

Substituting (C.9) into (C.2) and using

$$\int_{-\pi}^{\pi} \omega^2 e^{-j\omega l} d\omega = \begin{cases} \frac{2\pi^3}{3}, & l = 0, \\ \frac{4\pi(-1)^l}{l^2}, & l \neq 0, \end{cases} \quad (\text{C.10})$$

we get the criterion for selecting the interpolation factor I :

$$I > \frac{\sigma_d}{\sigma_T} \sqrt{\frac{\pi^2}{3} + \frac{2}{\sigma_d^2} \sum_{\substack{l=-\infty \\ l \neq 0}}^{\infty} \frac{(-1)^l}{l^2} E[d_r(n) d_r(n+l)]}, \quad (\text{C.11})$$

where $\sigma_d^2 \equiv E[d_r^2(n)]$ is the power in $d_r(n)$ and also that in $d(n)$.

Since the right-hand side of (C.11) depends very much on the statistics of $d_r(n)$, we now seek an upper bound as a worst-case requirement for I . The extreme case that maximizes the right-hand side of (C.11) is

$$E[d_r(n) d_r(n+l)] = \sigma_d^2 (-1)^l, \quad \forall l \in (-\infty, \infty), \quad (\text{C.12})$$

which, for example, corresponds to $d_r(n)$ being a sine wave of the Nyquist frequency. In this case,

$$\frac{1}{\sigma_d^2} \sum_{l=1}^{\infty} \frac{(-1)^l}{l^2} E[d_r(n) d_r(n+l)] = \sum_{l=1}^{\infty} \frac{1}{l^2} = \frac{\pi^2}{6}. \quad (\text{C.13})$$

The last step above is a mathematical formula.

Substituting (C.13) into (C.11), the worst-case requirement is found as

$$I > \pi \left(\frac{\sigma_d}{\sigma_T} \right) \equiv \pi \cdot 10^{\text{TR}/20}, \quad (\text{C.14})$$

where $\text{TR} \equiv 20 \log(\sigma_d/\sigma_T)$ is the target ratio (in dB) of the power of $d(n)$ to the tolerable power of the adjustment errors. For example, if one wants the adjustment errors to be at least 30 dB below $d(n)$ in level, then (C.14) suggests that

$$I > \pi \cdot 10^{30/20} \approx 99.3. \quad (\text{C.15})$$

This results in a choice of $I = 100$.

Acknowledgments

The Royal Canadian Mounted Police partially funded this research and provided test cases. The authors would also like to thank Dr. Bradford Gover, of the Institute for Research in Construction, National Research Council, for organizing the subjective evaluation and analyzing its results. Last but not least, the authors are grateful to the reviewers, whose invaluable comments and suggestions helped improve the paper.

References

- [1] B. E. Koenig, D. S. Lacey, and S. A. Killion, "Forensic enhancement of digital audio recordings," *Journal of the Audio Engineering Society*, vol. 55, no. 5, pp. 352–371, 2007.
- [2] A. H. Sayed, *Fundamentals of Adaptive Filtering*, John Wiley & Sons, New York, NY, USA, 2003.
- [3] Tandy Corporation, "Owner's Manual: DUoFONE® TAD-320 Tone Remote Control Telephone Answering System with Voice Synthesized Time and Date," 1986.
- [4] Q. Li, C. He, and W.-G. Chen, "Challenges and solutions for designing software AEC on personal computers," in *Proceedings of the 11th International Workshop for Acoustic Echo and Noise Control (IWAENC '08)*, Seattle, Wash, USA, September 2008.
- [5] M. Pawig, G. Enzner, and P. Vary, "Adaptive sampling rate correction for acoustic echo control in voice-over-IP," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 189–199, 2010.
- [6] T. F. Quatieri and G. C. O'Leary, "Far-echo cancellation in the presence of frequency offset," *IEEE Transactions on Communications*, vol. 37, no. 6, pp. 635–644, 1989.
- [7] D. G. Messerschmitt, "Asynchronous and timing jitter insensitive data echo cancellation," *IEEE Transactions on Communications*, vol. 34, no. 12, pp. 1209–1217, 1986.
- [8] H. Ding and D. I. Havelock, "Drift-compensated adaptive filtering to improve speech intelligibility in presence of asynchronous interference," in *Proceedings of the 16th International Conference on Digital Signal Processing (DSP '09)*, Santorini, Greece, July 2009.

- [9] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice Hall, Upper Saddle River, NJ, USA, 3rd edition, 1996.
- [10] H. Ding, "Fast affine projection adaptation algorithms with stable and robust symmetric linear system solvers," *IEEE Transactions on Signal Processing*, vol. 55, no. 5 I, pp. 1730–1740, 2007.
- [11] H. Ding, Y. Chu, and X. Qiu, "Voice separation using ratchet FAP algorithm," in *Proceedings of the Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*, pp. 57–60, Trento, Italy, May 2008.
- [12] H. Ding, "Detecting instability potentials in regularization for fast affine projection algorithms," in *Proceedings of the 41st Asilomar Conference on Signals, Systems and Computers (ACSSC '07)*, pp. 1623–1627, Pacific Grove, Calif, USA, November 2007.
- [13] IEEE, "IEEE recommended practice for speech quality measurements," *IEEE Trans Audio Electroacoust*, vol. 17, no. 3, pp. 225–246, 1969.