

Coherent Time-Domain Canceling of Interference for Radio Astronomy

S.W. ELLINGSON¹ AND R.M. BUEHRER¹

¹*Bradley Dept. of Electrical & Computer Engineering
Virginia Tech
Blacksburg, VA 24061, USA*

ABSTRACT

Radio astronomy is vulnerable to interference from a variety of anthropogenic sources. Among the many strategies for mitigation of this interference is coherent time-domain canceling (CTC), which ideally allows one to “look through” interference, as opposed to avoiding the interference or deleting the afflicted data. However, CTC is difficult to implement, not well understood, and at present this strategy is not in regular use at any major radio telescope. This paper presents a review of CTC including a new comprehensive study of the capabilities and limitations of CTC using metrics relevant to radio astronomy, including fraction of interference power removed and increase in noise. This work is motivated by the emergence of a new generation of communications systems which pose a significantly increased threat to radio astronomy and which may overwhelm mitigation methods now in place.

Keywords: instrumentation : detectors – methods : analytical

1. INTRODUCTION

Interference of anthropogenic origin is an old but growing problem for radio astronomy. While certain frequency bands are in some sense set aside for exclusively passive uses, less than 2.1% of the spectrum below 3 GHz is protected in this manner (National Research Council 2010). In fact, most of the spectrum that is necessary and commonly used for radio astronomy is in frequency bands in which radio astronomy receives little or no regulatory protection. Astronomy in bands not explicitly protected for radio astronomy is possible only because large swaths of the time-frequency plane remain fallow, and because astronomers have become expert at editing data in order to remove interference from sparsely-used regions of the time-frequency plane. An overview of these techniques appears in ITU Radiocommunication Bureau (2013a).

By far, the most commonly-used category of interference mitigation techniques consists of detecting time-frequency pixels that are corrupted by interference, and then eliminating those pixels from subsequent processing, typically after the observation. In this paper, we refer to this as *incoherent time-frequency editing* (ITFE).

ITFE is effective because modern instruments typically reduce Nyquist-rate time-domain signals to time-frequency “dynamic spectrum” representations having resolutions ranging from microseconds to seconds in the time domain, and kHz to MHz in the frequency domain. This is necessary in order to accommodate the limited bandwidth and capacity of modern data storage systems. This intermediate form of the data is useful for identification of interference and provides a convenient opportunity to excise the affected time-frequency pixels prior to a subsequent reduction to science products such as averaged spectrum for spectroscopy, and dedispersed and averaged pulse profiles for pulsar processing. ITFE algorithms have been refined and fine-tuned over time, culminating in sophisticated and highly-effective software such as “flagdata” in the interferometer data analysis software CASA,¹ “rfifind” in the pulsar analysis software PRESTO,² and “AOFlagger” (Offringa et al. 2012).

This state of affairs may not be sustainable. Strong societal, economic, and political pressures exist to increase the utilization of spectrum, including in remote areas where radio telescopes tend to be located. A particularly ominous development in this regard is the dramatic

Corresponding author: S.W. Ellingson
ellingson@vt.edu

¹ <https://ascl.net/1107.013>

² <https://ascl.net/1107.017>

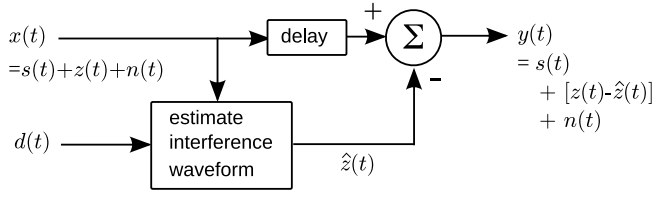


Figure 1. Coherent time-domain canceling (CTC), feed-forward architecture. The purpose of the “delay” block is to accommodate the latency of the “estimate interference waveform” block.

increase in the use of satellites to deliver world-wide continuous broadband communications. Whereas previous generations of systems consisted of a few satellites in geosynchronous orbit (e.g., INMARSAT), or tens of satellites in low-earth orbit (LEO; e.g., Iridium), emerging and planned systems consist of tens of thousands of satellites in LEO, transmitting in L-band and X-band (European Conference of Postal and Telecommunications Administrations (CEPT) 2019; Kodheli et al. 2020; United Nations Office of Outer Space Affairs and the International Astronomical Union 2021). Soon there will be no location on Earth which is not within view of many such satellites simultaneously. Interference from terrestrial communications is also expected to worsen with the deployment of new generations of wireless communications systems and navigation and positioning systems using radio frequencies. Compounding the problem is the fact that future generations of radio telescopes will consist of 100s of antennas deployed over areas 100s of km in extent, and will therefore will be geographically commingled with interference sources that previously could be avoided simply by siting in remote locations. Therefore, it is uncertain whether ITFE will continue to be sufficient; at some point the amount of data that must be excised renders the remainder unsuitable for scientific interpretation; and even if this is not the case, the still-formidable amount of manual effort required to process data using ITFE may become intractable.

One possible solution lies in spatial processing. Telescopes with array feeds, or telescopes which are themselves arrays, have in principle the ability to form pattern nulls in the directions from which interference arrives. While this strategy has been well-studied, it is not in regular use in any major radio telescope. Reasons include (1) high system cost/complexity and (2) undesirable dynamic modification of main lobe gain and overall pattern characteristics which are difficult to know or correct in subsequent processing.

An alternative strategy, and the topic of this paper, is coherent time-domain canceling (CTC), illustrated in Figure 1. (The particular form shown in this figure is the “feedforward” architecture. An alternative “feedback”

architecture is shown in Figure 8 (Section 6)). In Figure 1, the signal $x(t)$ from the instrument is the sum of the astronomical signal of interest (SOI) $s(t)$, interfering signal $z(t)$, and noise $n(t)$. The signal $x(t)$ is compared to a “reference signal” $d(t)$ which represents the best available information about $z(t)$. The reference signal may be obtained either from an external input, such as a separate antenna pointed at the source of the interference; or internally synthesized; e.g., based on *a priori* information about $z(t)$. The result of the comparison is used to create the interference estimate $\hat{z}(t)$, which is subsequently subtracted from $x(t)$, yielding the output $y(t) = s(t) + [z(t) - \hat{z}(t)] + n(t)$. Ideally, this operation completely removes the interference (i.e., $z(t) - \hat{z}(t) = 0$) while preserving $s(t)$ and (importantly in radio astronomy) $n(t)$ with negligible distortion. Thus, CTC potentially allows an instrument to “look through” interference and, unlike spatial processing, is applicable also to single-feed instruments and instruments employing fixed analog beamforming, such as certain kinds of focal plane arrays and radio cameras. Note that the “look through” capability is not merely deleting interference, but (unlike ITFE) is potentially restoring the use of the afflicted spectrum for astronomy.

Despite these compelling features, and like spatial processing, no major radio telescope regularly employs CTC. The reasons are somewhat similar: Increased system cost/complexity, and the potential for increased noise and signal distortion that may be difficult to know or correct in subsequent processing.

The purpose of this paper is to provide a review of CTC for radio astronomy, provide new information about capabilities and limitations, and provide a new starting point for those interested in revisiting this technology. This paper is organized as follows. Section 2 addresses the important preliminary question of how effective CTC needs to be in order to achieve the desired “look through” capability; and also the distinction between CTC for radio astronomy and CTC for communications, radar, and other active radio frequency applications. Section 3 presents the theory of optimal CTC design, and what constitutes “optimal” in this application. In Section 4 we provide a new and comprehensive analysis of the performance of optimal canceling including an example using real-world data. Section 5 presents a canceler with reduced complexity, but similar performance. Whereas Sections 3 through 5 address the “feedforward” architecture depicted in Figure 1, Section 6 addresses the alternative “feedback” architecture, which exhibits similar performance in certain conditions, but which may be less well-suited to radio astronomy. Section 7 addresses practical considerations that apply

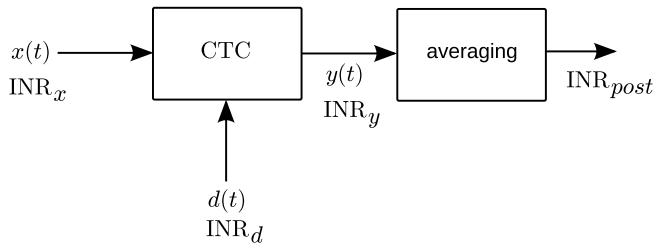


Figure 2. System model for analysis of the performance of CTC for radio astronomy.

to the implementation of CTC in radio astronomy. Section 8 presents a brief review of past work on CTC for radio astronomy. We have made the unconventional choice of presenting this review at the end so that past work can be understood in the context of the theory and concepts presented in this paper.

2. HOW MUCH CANCELING IS REQUIRED?

A fundamental difference between CTC and ITFE is that CTC cannot *completely* remove interference. Whereas ITFE removes 100% of the interference that is detected, CTC is limited by estimation error even if the interference is reliably detected. This raises the question of how much canceling is required, which in turn raises the question of how much interference is detrimental. The answers depend on the application: See [National Research Council \(2010\)](#) for a general overview and [ITU Radiocommunication Bureau \(2003, 2015\)](#) for levels that have traditionally been considered detrimental to radio astronomy. What follows is a generic analysis that provides context for the performance levels reported later in this paper.

Consider the system model shown in Figure 2. Here, INR_x is the interference-to-noise ratio (INR) at the input of the canceler, INR_y is the INR at the output of the canceler, and INR_{post} is the INR following whatever averaging is subsequently applied. For simplicity and with no loss of generality, let us assume INR_x , INR_y , and INR_{post} are each evaluated for the same bandwidth B . To quantify the amount of canceling, let us define “interference rejection ratio” (IRR) to be the ratio of the time-average power of interference in the input to time-average power of interference at the output. (This definition is formalized in Section 3. For the purposes of this section, the definition as stated suffices.) Note $\text{IRR} = 1$ for no canceling and $\text{IRR} \rightarrow \infty$ with improving performance.

Averaging increases the SOI signal-to-noise ratio as well as INR_{post} in proportion to $\sqrt{B\Delta t}$, where Δt is the averaging time. Normally Δt is selected to make the SOI signal-to-noise ratio $\gg 1$, whereas INR_{post} is ideally

$\ll 1$ so as to have negligible effect on the observation. Therefore the amount of canceling required to effectively mitigate an interferer, assuming the noise is unaffected by the canceler, is

$$\text{IRR} \gg \text{INR}_x \cdot \sqrt{B\Delta t}, \quad (1)$$

and this is necessary even if $\text{INR}_x \ll 1$. It will be useful later in this paper to have this condition in the form of a specific numerical threshold that can be compared to results. For this purpose we define

$$\text{IRR}_{\text{req}} = 10 \cdot \text{INR}_x \cdot \sqrt{B\Delta t} \quad (2)$$

where the constant 10 is arbitrary but reasonable in light of the preceding discussion.

To clearly see the implications, consider an observation with $\sqrt{B\Delta t} = 100$; for example, $B = 10$ kHz and $\Delta t = 1$ s. First, a strong interferer appears, having $\text{INR}_x = 10^3$. Without CTC, $\text{INR}_{\text{post}} = 10^5$ after averaging; thus $\text{IRR}_{\text{req}} = 10^6$ (60 dB). As will be demonstrated in Section 4, this is on the high end of plausible values of IRR and requires that INR_d , the interference-to-noise ratio in the reference channel $d(t)$, be very high. This level of performance also requires that no implementation issues (addressed in Section 7) significantly degrade IRR. It should also be noted that this is a regime which has been well-explored in the literature on communications, radar, and other active radio frequency systems (see e.g. [Ghose \(1996\)](#)).

However an even more challenging scenario emerges when averaging converts weak interference into strong interference. Continuing the example: An interferer having $\text{INR}_x = 10^{-1}$ emerges with $\text{INR}_{\text{post}} = 10$ without CTC, and so becomes detrimental despite being very weak. Here, $\text{IRR}_{\text{req}} = 10^2$ (20 dB). Although this IRR is relatively modest, it must be achieved for a much lower INR_x , and perhaps also with a much lower INR_d . As we shall see later in this paper, low INR_d also increases the risk that significant additional noise is injected into the output. Thus, ironically, this weak interference may be more difficult to mitigate than the strong interference considered in the previous paragraph. This is a regime which has *not* been well-explored in the communications, radar, and navigation literature because INR_{post} is typically not much greater than INR_y in these applications, and furthermore there is typically no advantage in driving INR_y or INR_{post} below 1 in these applications. For these reasons, CTC techniques which are effective for communications, navigation, and radar are not necessarily suitable for radio astronomy.

3. OPTIMAL TIME-DOMAIN CANCELING

3.1. Derivation & Implementation

We now describe the optimal implementation of the “estimate interference waveform” block in Figure 1, which somehow computes the estimate $\hat{z}(t)$ using $x(t)$ and $d(t)$. A broad class of relevant applications is addressed by assuming $d(t)$ has the form

$$d(t) = f(\tau) * s(t) + g(\tau) * z(t) + u(t) \quad (3)$$

where $f(\tau)$ and $g(\tau)$ are impulse responses describing the difference between how $s(t)$ and $z(t)$, respectively, appear in the reference channel relative to the system input, $u(t)$ is the noise in the reference channel, and “*” denotes convolution. This suggests implementation of the “estimate interference waveform” block as a filter having impulse response $h(\tau)$; i.e.,

$$\hat{z}(t) = h(\tau) * d(t) \quad (4)$$

To determine $h(\tau)$, we first note

$$\hat{z}(t) = h(\tau) * f(\tau) * s(t) + h(\tau) * g(\tau) * z(t) + h(\tau) * u(t) \quad (5)$$

so ideally $h(\tau) * f(\tau) = 0$, $h(\tau) * g(\tau) = 1$, and the time-average power associated with the third term (noise) is minimized.

A general solution meeting these criteria is not possible because $f(\tau)$ and $g(\tau)$ are not precisely known *a priori*. To make progress, we assume that the time-average power associated with the first term is much less than the time average power associated with the second term; i.e.,

$$\langle |h(\tau) * f(\tau) * s(t)|^2 \rangle \ll \langle |h(\tau) * g(\tau) * z(t)|^2 \rangle \quad (6)$$

where the angle brackets denote mean over time. This condition is not hard to meet since the magnitude of $f(\tau) * s(t)$ can be made sufficiently small compared to that of $g(\tau) * z(t)$ in a properly-designed canceling system. For example, if $d(t)$ is obtained using a separate antenna or beam, that antenna or beam would be designed to have low gain in the direction of the SOI and relatively high gain in the direction of the interference. For the parametric estimation and subtraction (PES) strategy described in Sec. 8, $d(t)$ is generated internally and so for these methods $f(\tau) * s(t)$ is effectively zero.

Assuming Equation 6 applies, it is possible to design $h(\tau)$ to minimize the mean square error (MSE) defined as follows:

$$\text{MSE} = \langle |x(t - t_p) - h(\tau) * d(t)|^2 \rangle \quad (7)$$

where t_p is the delay indicated in Figure 1, and is now seen to be the “pipeline delay” associated with filtering.

Although minimizing MSE is not necessarily equivalent to forcing $h(\tau) * g(\tau) = 1$, minimizing MSE does maximize the interference-to-noise ratio in $\hat{z}(t)$, and is in this sense optimal.

At this point it is convenient to switch to discrete time notation. Let $\mathbf{d}[k]$ be M consecutive samples of $d(t)$ organized as an $M \times 1$ vector as follows:

$$\mathbf{d}[k] = [d((k - M + 1)T_S) \ d((k - M + 2)T_S) \ \dots \ d(kT_S)]^T \quad (8)$$

where T_S is the sample period and k is an integer. Also, we define \mathbf{w}^* (“*” denoting the conjugate) to be the $M \times 1$ vector representing $h(\tau)$.³ Equation 7 may now be written in discrete complex baseband form as follows:

$$\text{MSE} = \langle |x(kT_S - t_p) - \mathbf{w}^H \mathbf{d}[k]|^2 \rangle \quad (9)$$

where “ H ” denotes the conjugate transpose and “ $\langle \cdot \rangle$ ” now operates over k .

It is well known (see e.g. Haykin (2001)) that the filter \mathbf{w} which minimizes MSE is the solution to

$$\mathbf{R} \mathbf{w} = \mathbf{r} \quad (10)$$

where \mathbf{R} is the $M \times M$ covariance matrix

$$\mathbf{R} = \langle \mathbf{d}[k] \mathbf{d}^H[k] \rangle \quad (11)$$

and \mathbf{r} is the $M \times 1$ reference correlation vector

$$\mathbf{r} = \langle x^*(kT_S) \mathbf{d}[k] \rangle \quad (12)$$

Finally, the filter output is

$$\hat{z}(kT_S) = \mathbf{w}^H \mathbf{d}[k] \quad (13)$$

This method is commonly known as “minimum MSE” (MMSE), and we refer to this specific implementation as “feedforward MMSE.”

There are three important things to know about feedforward MMSE in this application. First: To the extent that the inequality in Equation 6 is not satisfied, \mathbf{w} will be biased and the canceling of $z(t)$ will be degraded. Second: The same problem will result in the term $h(\tau) * f(\tau) * s(t)$ being non-zero in $\hat{z}(t)$, which will distort the SOI in the output of the canceler. Third: The term $h(\tau) * u(t)$ will be injected into the output of the canceler, which will decrease INR_y and color the noise in $y(t)$, so it is important that INR_d be as large as possible. The second and third items are aspects of what we refer to as “toxicity,” and are particularly important considerations for radio astronomy. This is because achieving

³ The use of \mathbf{w}^* as opposed to \mathbf{w} is arbitrary, but is customary and simplifies notation later.

the necessary IRR may be for naught if the SOI $s(t)$ or the primary channel noise $n(t)$ are distorted in a manner that impedes scientific interpretation. The toxicity issue is addressed further in Section 7.3.

To implement MMSE one must choose (1) the number of samples L used for “training;” i.e., used to compute \mathbf{R} and \mathbf{r} ; and (2) the filter length in samples, M . The training length L determines the accuracy to which \mathbf{w} is computed, which normally improves with increasing L . Thus, IRR normally increases with L . However L should be small enough that the change in the impulse response $g(\tau)$ is negligible relative to the time LT_S over which the canceler attempts to determine \mathbf{w} .

The filter length M also entails a tradeoff. The filter must be long enough to equalize the frequency response corresponding to $g(\tau)$ with sufficient accuracy. However, increasing M increases the effective duration of $h(\tau)$, which limits the ability of the filter to adapt to changing conditions. Thus, M should be small enough that the change in $g(\tau)$ is negligible relative to the time MT_S required for the filter to produce an output. Making M larger than is required to equalize the interference component of the reference signal may decrease IRR and is not recommended; see e.g. Table 1.

Finally, note that L should be $\gg M$ to ensure that \mathbf{R} is numerically well-conditioned (i.e., not nearly singular) and to ensure a low-variance estimate of \mathbf{r} .

3.2. Theoretical Performance

A complete rigorous derivation of the theoretical performance of the feedforward MMSE canceler is, to the best of our knowledge, not available. In Section A.1 we derive expressions for performance for the special case of $M = 1$. These are Equations A29–A31 and A34–A36. These expressions are validated by comparison to the simulation results in Section 4 (Figures 3, 4, and 5 and associated text), where the agreement is found to be excellent. Derivation of expressions for IRR for $M > 1$ is much more difficult and has not been completed. However Section A.2 presents empirical expressions for $M > 1$ (Equations A37–A42) which are again shown to be in excellent agreement with the simulation results.

This paper considers two similar but distinct definitions of IRR. “IRR₁” is defined as the ratio of time-average power of the interference in the input to time-average power of the interference in the output; i.e.,

$$\text{IRR}_1 = \frac{\langle |z(t)|^2 \rangle}{\langle |z(t) - h(\tau) * g(\tau) * z(t)|^2 \rangle} \quad (14)$$

This is arguably the “natural” definition of IRR. However this definition does not account for noise injected by

the canceler into the output that could be interpreted as new interference. Furthermore, this metric may be difficult to measure experimentally. Therefore we define an alternative metric “IRR₂” to be the ratio of time-average power of the interference in the input to time-average power of the difference between $z(t)$ and the interference estimate $\hat{z}(t)$ in the output; i.e.,

$$\text{IRR}_2 = \frac{\langle |z(t)|^2 \rangle}{\langle |z(t) - \hat{z}(t)|^2 \rangle} \quad (15)$$

As noted in Appendix A and demonstrated in the results presented in the following sections, IRR₁ and IRR₂ are usually equal when INR_d is large, but are significantly different otherwise. Our impression is that IRR₂ is probably most appropriate where the spectrum of the output is less important than the total power of the output; e.g., continuum and most pulsar observations. On the other hand, IRR₁ is perhaps more appropriate if the spectrum is the primary concern – in particular, in spectroscopy – since IRR₁ does not conflate canceler noise injection with interference suppression.

4. HOW MUCH CANCELING IS POSSIBLE?

In this section we quantify the performance of feedforward MMSE CTC using a combination of simulations, derived expressions, empirical expressions, and an example using real-world data.

4.1. Experiment Design

In each simulation, the interference consists of a single signal which is either a sinusoid or zero-mean white Gaussian noise. The sinusoidal interference waveform is representative of interference which is narrowband in the sense that the bandwidth of $z(t)$ cannot be spectrally resolved. When the interference is noise, it fills the Nyquist bandwidth, and can be viewed as the limiting case where the bandwidth of $z(t)$ exceeds the bandwidth of the observation. When the interference is sinusoidal, the frequency is varied from trial to trial according to a uniform random distribution from $-\pi/2$ to $+\pi/2$ radians/sample. The primary-to-reference channel response for the SOI, $f(\tau)$, is zero; i.e., there is no astronomy ingress into the reference channel. The primary-to-reference channel response for the interference, $g(\tau)$, is a constant with magnitude determined by the specified INR_d and with phase varied from trial to trial according to a uniform random distribution from $-\pi$ to $+\pi$ radians. The primary and reference channel noise waveforms ($n(t)$ and $u(t)$, respectively) are uncorrelated zero-mean white Gaussian noise, and $n(t)$ and $u(t)$ are uncorrelated with $z(t)$ in scenarios where $z(t)$ is a noise waveform.

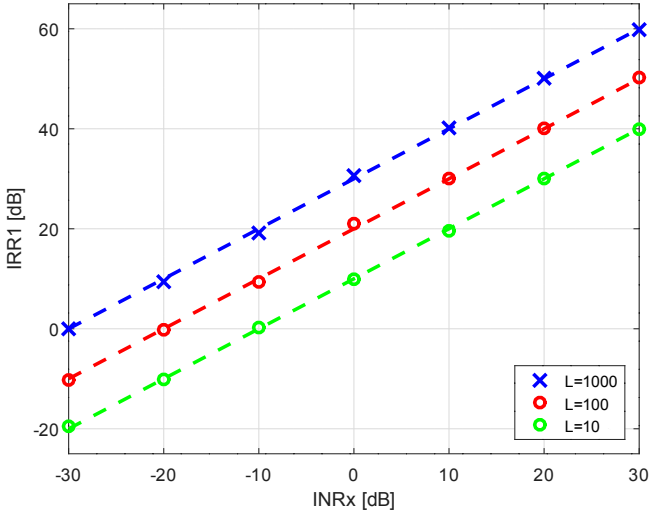


Figure 3. $\overline{\text{IRR}}_1$ for high INR_d (here, $\text{INR}_d = +70$ dB), $M = 1$. Markers are simulation results. Lines are theoretical results ($L \cdot \text{INR}_x$). Results are the same for sinusoidal and noise interference waveforms, and $\overline{\text{IRR}}_2$ is indistinguishable from $\overline{\text{IRR}}_1$ for either waveform.

In any given trial, IRR_1 and IRR_2 are computed over 10^6 samples. Statistics of IRR_1 and IRR_2 are computed over 100 trials. Care is required in computing these statistics. The mean of these quantities over trials is not an appropriate statistic, because IRR can be intermittently very high for a sinusoid having constant magnitude, phase, and frequency over the duration of the experiment.⁴ We solve this problem by reporting the mean over the trial values of the numerator of Equations 14 and 15 divided by the mean over trial values of the denominator of Equations 14 and 15.⁵ We refer to the statistics of IRR computed in this specific way as $\overline{\text{IRR}}_1$ and $\overline{\text{IRR}}_2$, respectively.

We also calculate noise ingress ratio (NIR), defined as the ratio of the time average power of $n(t) - h(\tau) * u(t)$, (the total noise in the output) to the time-average power of $n(t)$, again computed over 10^6 samples and averaged over 100 trials. The minimum and ideal value of NIR is 1 (0 dB), and a greater value indicates an increase in the effective system temperature.

4.2. High INR_d – Narrowband Interferer

⁴ This is especially important to know for hardware testing using synthesized interference signals.

⁵ This problem can also be avoided using median statistics, but the results will be slightly different, most notably in the high- INR_d regime. In this regime, the median over trials of IRR_1 is $\sqrt{2} \cdot \overline{\text{IRR}}_1$, and similarly the median over trials of IRR_2 is $\sqrt{2} \cdot \overline{\text{IRR}}_2$.

interferer	$\overline{\text{IRR}}_1$	$\overline{\text{IRR}}_2$
sinusoid	$L \cdot \text{INR}_x$	$L \cdot \text{INR}_x / M$
noise	$L \cdot \text{INR}_x / M$	$L \cdot \text{INR}_x / M$

Table 1. Performance in the high INR_d regime and $L \gg M$, summarized from results of simulations. $M = 1$ sinusoidal interferer results are verified by theory (Equations A30 and A35). $M > 1$ sinusoidal interferer results agree with the empirical equations A38 and A41.

We begin with the special case of sinusoidal interference and high INR_d . Figure 3 shows the results for $M = 1$, $\text{INR}_d = +70$ dB, varying INR_x and L . We find that the simulations are in excellent agreement with the analysis in Appendix A (Section A.1); that is: $\overline{\text{IRR}}_1 = \overline{\text{IRR}}_2 = L \cdot \text{INR}_x$. Note that IRR is proportional to both L and INR_x , even for $L \cdot \text{INR}_x < 1$.

Results for $M \geq 1$ are shown in the first row of Table 1. While it is not surprising that $\overline{\text{IRR}}_1$ is independent of M , the finding that $\overline{\text{IRR}}_2$ is inversely proportional to M is counter-intuitive.⁶ Clearly it is not safe to make M larger than necessary.

Using the definition from Section 2, IRR is judged to be sufficient if it is greater than IRR_{req} . Using the worst case from Table 1, $L \cdot \text{INR}_x / M \gtrsim 10 \cdot \text{INR}_x \cdot \sqrt{B\Delta t}$. Solving for L , we find

$$L \gtrsim 10\sqrt{B\Delta t} \cdot M \quad (16)$$

For example: Using the value of $\sqrt{B\Delta t} = 100$ from Section 2, we find $L \gtrsim 1000M$ is required for confidence that the interferer will be reduced to a negligible level in the output, and this does not depend on INR_x . Thus, one can plausibly achieve sufficient levels of canceling using feedforward MMSE when INR_d is high.

Now we consider NIR. For NIR we have only simulation results, but the findings are unambiguous. NIR does not depend on INR_x in this case. NIR *does* depend on the extent to which $L > M$, but this can easily be accommodated. For example: For $L = 1000$ and $M = 8$, NIR is merely 0.03 dB. NIR is decreased by increasing L or decreasing M , and is too small to be reliably measured when $L \geq 100$ and $M \leq 4$. Examples of high NIR due to *inappropriate* choices of L and M are NIR = 0.3 dB and 5 dB for $M = 8$ and $L = 100$ and 10, respectively. Summarizing: The aspect of toxicity measured by NIR is best managed by minimizing M and making $L \gg M$, and can be made negligible for reasonable values of M and L .

4.3. High INR_d – Wideband Interferer

⁶ This phenomenon is also apparent by comparing Figures 4 and 5.

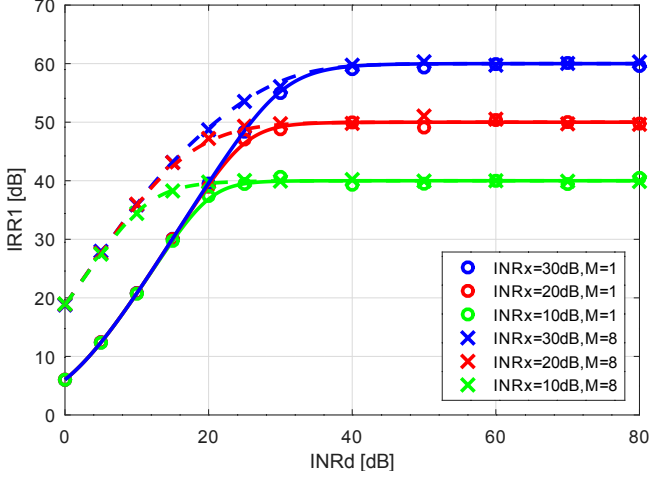


Figure 4. $\overline{\text{IRR}}_1$ as a function of INR_d . Sinusoidal interferer, $L = 1000$. Markers: Simulation. Solid lines: Theoretical equation for $M = 1$ (Equation A34). Dashed lines: Empirical equation for $M = 8$ (Equation A40).

The second row of Table 1 summarizes IRR for noise interference and high INR_d . It is not surprising that IRR is proportional to $L \cdot \text{INR}_x$, as in the case of sinusoidal interference (Section 4.2). However, in this case we find that *both* $\overline{\text{IRR}}_1$ and $\overline{\text{IRR}}_2$ are inversely proportional to M . The reason for this peculiar dependence on M is unclear and we continue to investigate.

In contrast to the sinusoidal interferer scenario, NIR in the wideband interferer scenario is always too small to measure reliably (here, < 0.01 dB), independent of M and L . The reason for the surprisingly good NIR performance in this case is that the canceler’s “estimate interference waveform” block converges to approximately flat magnitude response when the interference is spectrally-white noise, but is constrained only at one frequency – i.e., not necessarily flat and intermittently large – when the interference is sinusoidal. The latter facilitates increased injection of reference channel noise into the canceler output.

4.4. Reduced INR_d – Narrowband Interferer

Next, we consider the effect of reducing INR_d . Figures 4 and 5 show $\overline{\text{IRR}}_1$ and $\overline{\text{IRR}}_2$, respectively, for the sinusoidal interferer, varying INR_d , INR_x and M .

Considering first $M = 1$, note that the agreement between simulation and theory is excellent for both IRR metrics. As expected, the overall behavior depends on INR_d relative to $\text{INR}_x L$. The high INR_d regime is discussed in Section 4.2. For the low INR_d regime, the results are summarized in the first row of Table 2. Note that in this regime, IRR depends only on INR_d and M , but not on INR_x , and not on L as long as $L \gg M$. The

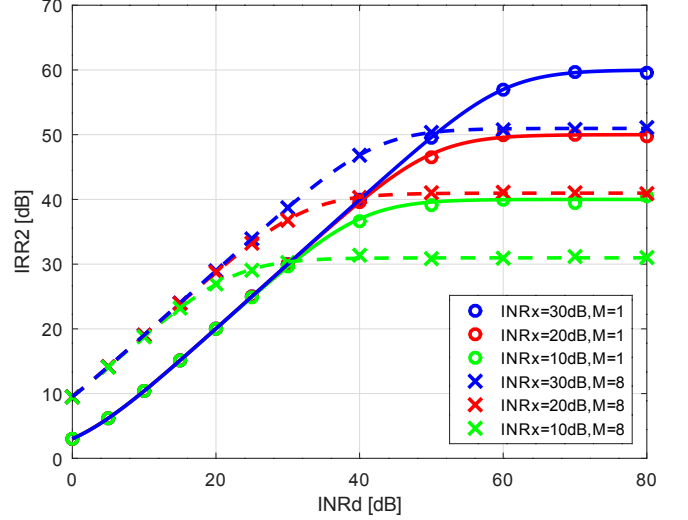


Figure 5. $\overline{\text{IRR}}_2$ as a function of INR_d . Sinusoidal interferer, $L = 1000$. Markers: Simulation. Solid lines: Theoretical equation for $M = 1$ (Equation A29). Dashed lines: Empirical equation for $M = 8$ (Equation A37).

interferer	$\overline{\text{IRR}}_1$	$\overline{\text{IRR}}_2$
sinusoid	$(M \cdot \text{INR}_d + 1)^2$	$M \cdot \text{INR}_d + 1$
noise	$(\text{INR}_d + 1)^2$	$\text{INR}_d + 1$

Table 2. Performance in the low INR_d regime and $L \gg M$, summarized from results of simulations. $M = 1$ sinusoidal interferer results are verified by theory (Equations A31 and A36). $M > 1$ sinusoidal interferer results agree with the empirical equations A39 and A42.

reason for the difference in dependence on INR_d between $\overline{\text{IRR}}_1$ and $\overline{\text{IRR}}_2$ is simply that the latter considers noise injected by the canceler to be interference, whereas the former does not. It is interesting to note that increasing M in the low- INR_d regime is beneficial, whereas this was found to be detrimental in the high- INR_d regime. The fact that IRR improves with increasing M in the low- INR_d regime indicates that the estimation filter is exhibiting spectral selectivity in this case.

Repeating the procedure in Section 4.2, we judge the canceling is sufficient if $(M \cdot \text{INR}_d + 1)^n \gtrsim \text{IRR}_{\text{req}}$, where $n = 2$ for $\overline{\text{IRR}}_1$ and $n = 1$ for $\overline{\text{IRR}}_2$. Solving for INR_d , we find

$$\text{INR}_d \gtrsim 10^{1/n} \text{INR}_x^{1/n} (B\Delta t)^{1/2n} M^{-1} \quad (17)$$

Let us consider the implications for $\overline{\text{IRR}}_1$ ($n = 2$). Using the value of $\sqrt{B\Delta t} = 100$ from Section 2, $\text{INR}_x = 10$ dB, and $M = 1$, we find $\text{INR}_d \gtrsim 20$ dB is required to have high confidence that the interferer will be reduced to a negligible level in the output. While this seems encouraging at first glance, consider what is required for

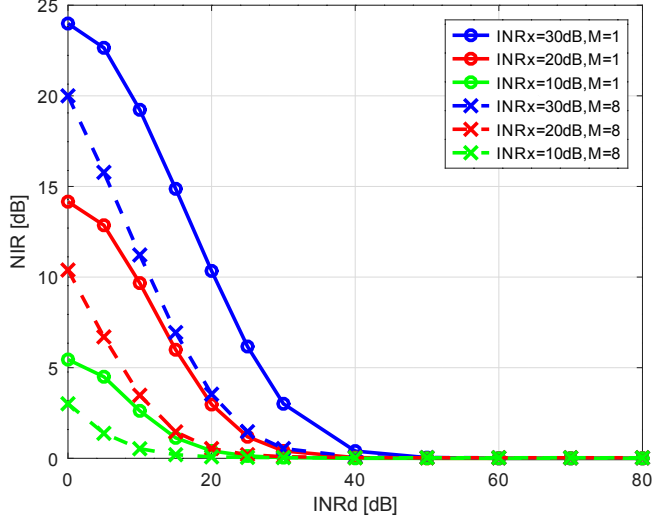


Figure 6. NIR as function of INR_d . Sinusoidal interferer, $L = 1000$. Markers: Simulation. Lines connect markers. The $M = 1$ curves also apply to the noise interferer, regardless of the actual value of M .

a weak interferer: For $\text{INR}_x = -10$ dB, $\text{INR}_d \gtrsim 10$ dB is required. While this value of INR_d is much lower, it must be achieved for an interferer which is much weaker. Specifically, the required ratio $\text{INR}_d/\text{INR}_x$ has increased from 10 dB to 20 dB. This does not bode well for CTC implementations in which the reference signal $d(t)$ is obtained from an auxiliary antenna.

Because INR_d is not necessarily high (as it was in Section 4.2), the potential for NIR to be significant is much greater. Figure 6 shows the situation for the sinusoidal interferer. Note that NIR can be devastatingly large for $\text{INR}_d < 40$ dB or so. Also note that the NIR catastrophe can be forestalled somewhat by increasing M .

4.5. Reduced INR_d – Wideband Interferer

IRR for the noise interferer in the low INR_d regime is summarized in the second row of Table 2. The single difference is that IRR does not depend on M , which is expected since the estimation filter is unable to exhibit spectral selectivity in this case.

As noted in Figure 6, the NIR performance for the noise interferer is the same as that for the sinusoidal interferer, except NIR does not depend on M . Again this attributable to the inability of the estimation filter to exhibit spectral selectivity in this case.

Before moving on, recall that the impulse response $g(\tau)$ for results in Section 4 is a complex valued constant and therefore represents a flat frequency response. To the extent that $g(\tau)$ represents a non-flat response and the resulting variation is significant over the spectrum of $z(t)$, M must necessarily be increased.

4.6. Real-World Example

In Appendix B we provide an example of the use of $M = 1$ feedforward MMSE to cancel a *bona fide* interference signal in a scenario representative of a typical radio astronomical observation. The interferer is an analog frequency modulation broadcast signal with bandwidth that dynamically varies from near zero to nearly the full bandwidth of the channel. The results are consistent with the results of the preceding sections, which confirms that the performance of $M = 1$ feedforward MMSE is not sensitive to the details of the interference waveform. Further, this example demonstrates good performance even in a case where $g(\tau)$ is demonstrably non-stationary.

5. REDUCED-COMPLEXITY FEEDFORWARD CANCELER

In the MMSE approach of Sections 3 and 4, the filter \mathbf{w} is the solution to $\mathbf{R}\mathbf{w} = \mathbf{r}$ (Equation 10). A simplified approach may be necessary or desirable. As we shall see in Section 5.2, simplifying the canceler does not necessarily result in a significant performance reduction.

5.1. Description

First, note that the covariance matrix \mathbf{R} depends only on the reference signal $d(t)$ and not at all on the input $x(t)$. So, we replace \mathbf{R} with a matrix that describes in some sense the time-average power of $d(t)$, but which facilitates a simple solution for \mathbf{w} . Such a matrix is $\|\mathbf{R}\|_2 \mathbf{I}$, where \mathbf{I} is the identity matrix and $\|\mathbf{R}\|_2$ is the induced 2-norm (largest singular value) of \mathbf{R} . A variety of computationally-efficient algorithms exist for accurate estimation of the largest singular value of a covariance matrix directly from samples (i.e., $d(kT_S)$ for a set of values of k). Subspace tracking (see e.g., DeGroat et al. (2010) and in particular Yang (1995)) is well-suited to this task. The solution of $\mathbf{R}\mathbf{w} = \mathbf{r}$ with this simplification is:

$$\mathbf{w} = \mathbf{r} / \|\mathbf{R}\|_2 \quad (18)$$

Note that this approach will entail some important disadvantages with respect to MMSE. First: Performance will be degraded if M is greater than 1 and the signal subspace of \mathbf{R} has rank greater than 1; i.e., has more than one significant singular value. Thus, degradation is expected if the interference has significant fractional bandwidth. A full-bandwidth noise interferer represents the worst case in this respect, since in that scenario of the rank of the signal subspace of \mathbf{R} is M . Second: To the extent that $d(t)$ contains signals other than the intended interference component and noise, these will not be mitigated by the resulting filter and will pass through to the canceler output. In contrast, MMSE will,

interferer	$\overline{\text{IRR}}_1 = \overline{\text{IRR}}_2$
sinusoid, any M	$L \cdot \text{INR}_x$
noise, $M = 1$	$L \cdot \text{INR}_x$
noise, $M > 1$	$< L \cdot \text{INR}_x$ (see e.g. Fig. 7)

Table 3. IRR of the reduced complexity feedforward MMSE method in the high INR_d regime and $L \gg M$, summarized from simulations. (Compare to Table 1.)

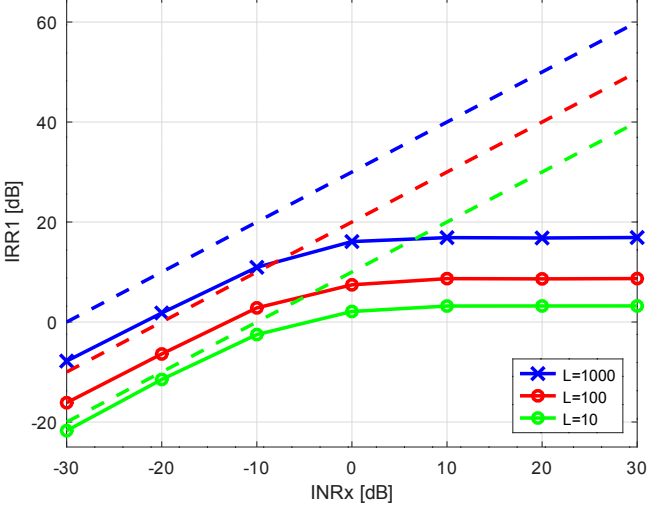


Figure 7. $\overline{\text{IRR}}_1$ for $\text{INR}_d = +70$ dB, $M = 8$, noise interference waveform. $\overline{\text{IRR}}_2$ is identical. Solid lines with markers: Reduced-complexity method; Dashed lines: MMSE.

to the extent that the degrees of freedom provided by M allow, attempt to mitigate these signals.

5.2. Performance

The experiments reported in Section 4 were repeated using Equation 18 (in lieu of MMSE) to generate \mathbf{w} . The results for the high INR_d regime are summarized in Table 3. Comparison to the results of the MMSE implementation (Table 1) reveals the following differences. First, and as expected, performance is degraded for the noise interferer when $M > 1$. An example is shown in Figure 7 ($M = 8$), which shows that IRR saturates at some threshold value of INR_x which decreases with increasing M .

Second, $\overline{\text{IRR}}_1 = \overline{\text{IRR}}_2$ in all cases considered. Specifically, $\overline{\text{IRR}}_2$ no longer depends on M . This is significant: If $\overline{\text{IRR}}_2$ is the metric that best describes performance in a particular application, and the interference is narrow-band, and $M > 1$, then the reduced complexity method actually *outperforms* MMSE. It is important to keep in mind, however, that $g(\tau)$ models a zero-length impulse response channel in these experiments; should the true impulse response have significant length such that

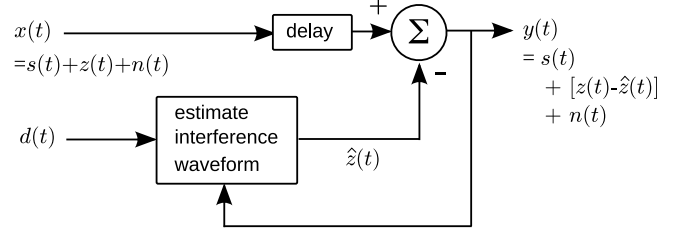


Figure 8. Feedback variant of the feedforward CTC canceler shown in Figure 1.

$M > 1$ is required for equalization, then this advantage of the reduced complexity method will be diminished.

In the low INR_d regime, the IRR performance of the reduced complexity method is the same as that of MMSE.

NIR performance is also somewhat different for the reduced complexity method relative to MMSE. In the high INR_d regime, NIR is always negligible. This is true even for sinusoidal interference in the $M > 1$ case; whereas for MMSE, NIR can become significant with increasing M . In the low INR_d regime, the NIR performance of the reduced complexity method is the same as that of MMSE.

Summarizing: The reduced-complexity method is probably an acceptable alternative to MMSE unless one of the following is true: (1) The interference has significant fractional bandwidth *and* M must be greater than 1; or (2) The reference channel $d(t)$ contains significant signals other than a single well-correlated version of $z(t)$, since these signals will not be mitigated in the reduced-complexity canceler as they are in the MMSE-based canceler, and therefore will be injected into the output at a significantly greater level. Item (2) is of particular concern for implementations in which $d(t)$ is obtained using a reference antenna.

6. FEEDBACK ARCHITECTURE

The CTC architecture shown in Figure 1 is “feedforward” because the interference input to the “estimate interference waveform” block is from the input of the canceler. The alternative is “feedback” architecture, shown in Figure 8, in which the interference input is from the *output* of the canceler. Although this architecture is not the primary topic of this paper, we address it here because it appears in several seminal papers on CTC in radio astronomy, notably Barnbaum & Bradley (1998) and Kesteven et al. (2005), and is also explored in Poulsen (2003), an interesting experiment at the Green Bank Telescope; so a comparison is warranted.

Before considering canceling performance, we identify two distinct and important characteristics of feedback architecture. First: The fact that \mathbf{w} depends on the output means that \mathbf{w} must be continuously updated during training. This is in contrast to feedforward architecture, where \mathbf{w} is determined at the end of a training period of length LT_S , and held constant until the end of the next training period. This may be either an advantage or disadvantage depending on the nature of the interference.

Second: Whereas the output of feedforward CTC is determined entirely by inputs (namely, $x(t)$ and $d(t)$), the output of feedback CTC depends also on past output. In signal processing terms, feedforward CTC has finite impulse response (FIR), whereas feedback CTC has infinite impulse response (IIR).

In this section we address specifically the least mean squares (LMS) algorithm (see e.g., Haykin (2001)), as it is relatively easy to study and is specifically the method used in Barnbaum & Bradley (1998) and Poulsen (2003). LMS is identical to feedforward MMSE with the exception that the filter is updated iteratively according to

$$\mathbf{w}((k+1)T_S) = \mathbf{w}(kT_S) + 2\mu y(kT_S)\mathbf{d}(kT_S) \quad (19)$$

The parameter μ controls the tradeoff between rapid convergence and agile tracking (requiring large μ) and low “jitter” following convergence (requiring small μ). It is a well-known rule-of-thumb that μ should be less than the reciprocal of the largest eigenvalue of \mathbf{R} ; i.e., $\mu < 1/(\text{INR}_d + 1)$ (Widrow et al. 1976). In practice, the optimal value of μ is typically not apparent without experimentation and tuning, and may of course also vary with circumstances. This is a disadvantage of LMS relative to feedforward architecture.

In the ideal (but unlikely) case that the jitter associated with μ is negligible, the IRR achieved by LMS after convergence is the same as that of feedforward MMSE. The effect of jitter is to degrade performance in the high INR_d regime. We have provided a derivation for $M = 1$ (analogous to the derivation provided for $M = 1$ feedforward MMSE) in Section A.3. In the high- INR_d regime,

$$\overline{\text{IRR}}_1 = \overline{\text{IRR}}_2 = \frac{1}{\mu} \frac{\text{INR}_x}{\text{INR}_d} \quad (20)$$

At first glance, the finding that IRR is inversely proportional to INR_d is counter-intuitive. The explanation for this is that the principal impairment in the high- INR_d regime is jitter, which is the net effect of the change in \mathbf{w} over the updates, and the magnitude of this change for any single update increases with increasing INR_d as is apparent from Equation 19.

Comparing Equation 20 to the corresponding feedforward MMSE result ($\text{INR}_x L$), we see that the IRR achieved by LMS compared to feedforward MMSE depends on $1/\mu \text{INR}_d$ relative to the number of training samples L used in feedforward MMSE. For example: At $\text{INR}_d = +70$ dB, LMS with $\mu \leq 10^{-9}$ would outperform feedforward MMSE with $L = 100$. On the other hand, decreasing μ comes at the expense of increasing convergence time for LMS, whereas increasing L comes at no analogous penalty for feedback MMSE, assuming $g(\tau)$ is stationary in both cases.

Returning to feedback architecture in general, it should be noted that the impact of \mathbf{w} jitter is not merely a reduction in IRR in the high INR_d regime. The jitter exists regardless of INR_d , and is potentially toxic for radio astronomy. Feedforward architecture, on the other hand, is not subject to \mathbf{w} -jitter, since in that architecture \mathbf{w} is obtained from a block of L samples and can be held utterly constant for as long as the scenario remains stationary.⁷

Finally, it should be noted that LMS is a “rank 1” algorithm in the same sense as the reduced complexity feedforward MMSE canceler of Section 5, and will have the associated limitations. While one might consider a MMSE implementation of the feedback architecture to address wideband interference, this has an extraordinarily greater computational burden relative to feedforward MMSE. This is because feedforward architecture requires correlation (Equations 11 and 12) only while training is in progress, and requires a solution to Equation 10 only when a new value of \mathbf{w} is needed. In contrast, the analogous implementation of feedback architecture estimates interference in the output, and therefore requires a new solution of Equation 10 for every sample processed.

7. PRACTICAL CONSIDERATIONS

In this section we address some particular issues that emerge in practical implementations of CTC.

7.1. Non-Stationarity Between the Primary and Reference Channels

MMSE-based CTC is potentially sensitive to the variations in the impulse response $g(\tau)$, defined in Equation 3, which describes the channel response applied to the interference signal in the reference channel relative

⁷ It is perhaps more accurate to say that feedforward architecture is vulnerable to \mathbf{w} -jitter, but over time scales of LT_S as opposed to T_S .

to the channel response applied to the interference signal in the primary channel.⁸

The derivation of MMSE-based CTC as well as the results presented in Sections 4–6 presume $g(\tau)$ to be perfectly stationary; i.e., independent of t . In feedforward CTC this means \mathbf{w} is assumed to be valid between updates, and in feedback CTC this means \mathbf{w} is assumed to be able to follow changes with negligible latency. This raises the question of the effect of non-stationarity on IRR.

In Appendix C we derive expressions for IRR for $M = 1$ feedforward MMSE, generalized from those in Section A.1. The non-stationarity is described as $g(\tau, t)$, which simplifies to $g(t)$ (i.e., a constant with respect to τ) for $M = 1$. It is found that the effect of time-varying $g(t)$ on IRR is negligible if

$$\epsilon^2 \ll (\text{IRR}|_{\epsilon=0})^{-1} \quad (21)$$

where ϵ^2 is the mean-squared variation of $g(t)$, and $\text{IRR}|_{\epsilon=0}$ is the associated IRR for stationary conditions. Thus, the impact of non-stationarity is greatest when IRR is high, decreases with decreasing IRR, and is negligible when Equation 21 is satisfied. For example, mean-square variation of 0.4 dB in the magnitude of $g(t)$ is significant if the IRR in stationary conditions would otherwise have been +40 dB, but is negligible if the IRR in stationary conditions is +20 dB.

Experiments using *bona fide* interference signals, such as those reported in Section 8 and Appendix B, provide evidence that non-stationarity exists but is not necessarily a show-stopper, especially if care is taken to use an appropriately short update rate. Nevertheless, potential adopters would be well-advised to carefully consider this issue in the design of CTC algorithms.

7.2. Intermittent Signals

While MMSE-based CTC is robust to the details of the interference waveform, there is a distinct and important type of waveform non-stationarity which can potentially cause problems: This is intermittency; i.e., signals which are not continuously present. One form of intermittency is burst modulation; examples being ground-based aviation radar and the Iridium user downlink.

Other forms of intermittency include temporarily-strong reflections from aircraft and interference from sources which transmit according some indiscernible schedule. CTC is certainly applicable in each of these cases; the problem is ensuring that the interference is present in the samples used to calculate the estimation filter.

Furthermore, it is preferable that the canceler operate only when the interference is present, and do nothing when the interferer is absent. This is an important consideration since a canceler operating in the absence of an interferer is prone to introduce spurious signals (more on this in Section 7.3). Thus, one encounters a problem in interference *detection*. Reliable interference detection is typically very difficult in the radio astronomy application since it is necessary to detect very weak interference as soon as it appears. In the case of burst modulations, individual bursts may not be present long enough to be reliably detected; see e.g. Ellingson & Hampson (2003) for an example where the performance of detection, and not the performance of CTC *per se*, limits overall IRR performance.

7.3. Toxicity

All forms of CTC entail adding the signal $\hat{z}(t)$ to the signal $x(t)$ from the telescope. Ideally $\hat{z}(t) = z(t)$, the interference component in $x(t)$. In practice, $\hat{z}(t)$ is the sum of (1) A waveform which is not quite equal to $z(t)$, (2) Spectrally-colored versions of signals that also appeared in $d(t)$ (noise and the astronomical SOI in particular), and (3) Internally-generated spurious content associated the operation of the canceler. The presence of these other signals in $\hat{z}(t)$ have a potentially deleterious effect on the processing and scientific interpretation of the data. This is what we refer to as “toxicity”. Three aspects of the toxicity problem already addressed include reference signal noise injection (quantified as NIR), \mathbf{w} -jitter, and spurious operation due to false detection (addressed in Section 7.2). Additional aspects of the toxicity problem include leakage of $s(t) + n(t)$ into the reference signal path, which is a problem particularly with auxiliary antennas; and spurious spectral content associated with block-wise updating of \mathbf{w} (see e.g. Ellingson (2020)).

7.4. Inadequate Reference Signal-to-Noise Ratio

A recurring theme in this paper has been the importance of a high-quality reference signal $d(t)$ with the highest possible INR_d . This poses a challenge in radio astronomy applications, since (as pointed out in Section 2) even interference which is much weaker than noise is potentially damaging. The solution employed in early studies of CTC for radio astronomy (see Section 8)

⁸ Since the interference waveform $z(t)$ appears in both the primary and reference channels, MMSE is not affected by the non-stationarity of $z(t)$ itself; e.g., by changes in carrier magnitude, carrier phase, and so on. It is also not affected by the non-stationarity of the propagation channels through which the interference waveform is received, as long as $g(\tau)$ remains constant. The problem emerges when $g(\tau)$ changes with time, and is a problem only because the process of filter synthesis in MMSE presumes this to be constant.

was to acquire the reference signal through a separate high-gain antenna (variously referred to as a “reference” or “auxiliary” antenna) directed at the source of the interference. This is certainly effective, but entails considerable additional complexity since the antenna must be pointed, and if the source is moving, the antenna must track accordingly. This is not only awkward to implement, but requires *a priori* or at least real-time knowledge of the presence and direction of sources. It should be noted that some existing and emerging radio telescope arrays employ architectures which provide multiple narrow steerable beams within the wider beam of a single element of the array. In principle these beams could be used in lieu of CTC auxiliary antennas, but only for interference which arrives from within the element pattern.

Another strategy for increasing INR_d is narrowband filtering with adaptive tuning. This scheme exploits the fact that interferers of interest often occupy only a small fraction of the bandwidth being processed. Thus, applying a relatively narrow filter at the center frequency of the interferer prior to the “estimate interference waveform” block in Figures 1 and 8 can dramatically increase INR_d , and has the additional benefit of excluding signals unrelated to the interference. Specifically, this technique excludes spectrally-disjoint portions of the astronomical SOI from the “estimate interference waveform” block, which provides further mitigation against toxicity. Note that essentially this scheme is employed in the example presented in Appendix B.

Yet another tool for improving INR_d and mitigating toxicity is parametric estimation and subtraction (PES), addressed in Section 8.

7.5. Nyquist-Rate Implementation

A barrier to adoption of CTC has been hardware implementation. Unlike ITFE, CTC requires access to a Nyquist-rate data stream. This presents a potential challenge in modern radio telescope implementations. Due to limitations in technology, storage cost, and logistics, existing instruments are typically limited to recording only the averaged spectrum. Therefore a practical operational CTC system must operate in real time, in the sense that any latency associated with CTC must be less than the time during which Nyquist-rate data is available. This in turn requires large amounts of high bandwidth memory and computing resources with low-latency access to this data. Thus, CTC is difficult to implement as an “add on” to existing instruments, and may require co-design and low-level integration with instrument electronics.

7.6. Separability

For the reasons cited in previous sections, it is not certain that CTC can be a fully “hands off, always on” capability for radio astronomy, and that astronomers will want the ability to enable, disable, or “tune” CTC as needed. Of course this is complicated by the issue noted in Section 7.5: CTC, unlike ITFE, must normally occur in real-time as the observation is running. Unless interference is certain to ruin an observation, astronomers might understandably prefer to keep CTC turned off, rather than to take the chance that data that might be salvageable using ITFE is instead ruined by CTC toxicity.

A possible remedy is *separability*, which might consist of any of the following techniques: (1) Record two versions of the observation: one with CTC, and the other without. (2) Record only the CTC-processed observation, but also $\hat{z}(t)$ so that it is possible to know precisely how CTC affected the data, and thereby retain some ability to perform remedial post-observation processing. This is feasible since the bandwidth of the interference is normally much less than the bandwidth of the observation. (3) Record only the observation without CTC, but also $\hat{z}(t)$. This retains the option to perform enhanced post-observation interference mitigation, although probably not truly coherent time-domain canceling. (4) If the observation can be recorded at the Nyquist rate, then full separability is possible simply by also recording $\hat{z}(t)$. (5) Record the observation at the Nyquist rate and also record $d(t)$ (as opposed to $\hat{z}(t)$), allowing full CTC to be implemented as a post-processing operation.

Options (4) and (5) have the benefit that CTC can be optimized after the observation, in the same manner as present-day ITFE processing.

8. A BRIEF HISTORY OF CTC IN RADIO ASTRONOMY

We now present a brief review of the history of CTC in radio astronomy. We have chosen not to attempt a numerical comparison between the findings of these studies and findings presented in this paper. This is partially due to the difficulty of extracting and presenting the relevant data from each paper in a consistent way, but also because experimental results are limited by practical factors in the implementation (typically well documented in the papers) that have a large effect on the outcomes. We strongly encourage readers to instead consult these papers directly; our paper may aid the reader by providing context.

Seminal work on canceling for radio astronomy appears in Barnbaum & Bradley (1998). This work addresses interference from radio stations in the 88–

108 MHz FM broadcast band. Their approach is feedback CTC using LMS with a reference signal obtained from a directional antenna pointed toward the source of the interference. They show theoretically that $\overline{\text{IRR}}_1 \approx (\text{INR}_d + 1)^2$, which is consistent with the low- INR_d regime result obtained in this paper (Equation A51). The authors present experimental results that are consistent with their theoretical analysis.

Ellingson (2002) reports experiments using feedforward MMSE to mitigate interference from the L-band satellite navigation system GLONASS in the main beam of a 3 m dish using the orthogonal linear polarization as a reference signal (thus, $\text{INR}_d = \text{INR}_x$; and precluding use for astronomy, since $f(\tau)$ is significant in this configuration). Results indicated $\text{IRR} > \text{IRR}_{req}$. Also, this work identifies the high- INR_d relationship $\text{IRR} = L \cdot \text{INR}_d$, obtained as a special case in this paper (Equations A30 and A35).

Poulsen (2003) reports experiments using LMS to mitigate GLONASS received through sidelobes of the Green Bank Telescope using a reference signal obtained from a separate 3.6-m reflector antenna tracking the interferer. $\text{IRR} > \text{IRR}_{req}$ is apparent despite challenges in setting the LMS step gain μ and mitigating non-stationarity in $g(\tau)$.

Kesteven et al. (2005) demonstrate that interference from a digital TV station at 675 MHz can be sufficiently suppressed to facilitate productive pulsar observations. Their work also employs feedback architecture with an auxiliary antenna, but they use a different method for computing \mathbf{w} that is similar to the reduced-complexity method of Section 5. They include theoretical analysis showing $\overline{\text{IRR}}_2 \approx \text{INR}_d + 1$, which again is consistent with the low- INR_d regime result obtained in this paper (Equation A48).

The fact that Barnbaum & Bradley (1998) perform analysis in terms of $\overline{\text{IRR}}_1$ and Kesteven et al. (2005) perform analysis in terms of $\overline{\text{IRR}}_2$ explains why these two similar techniques should yield such dramatically different IRR performance: We now see that the issue is simply that they used different performance metrics. Also, we note that neither work identifies the fact that their analysis is limited to the low INR_d regime, and that IRR in the high INR_d regime is significantly different, as explained in Section 6.

An important finding in these studies, and a recurring theme in this paper, is the need for large INR_d in order to effectively suppress weak interference. An approach that addresses this problem is *parametric estimation and subtraction* (PES). PES takes advantage of the fact that essentially all communications, radar, and navigation signals are comprised of modulated si-

nusoidal carriers which can be modeled as

$$z(t) = A(t) \cos[\omega_c t + \omega_\Delta(t) \cdot t + \theta(t)] \quad (22)$$

where $A(t)$, $\omega_\Delta(t)$, and $\theta(t)$ are parameters that vary slowly relative to the period of the carrier $2\pi/\omega_c$. This makes it possible to estimate these parameters; in fact, the process of estimating these parameters is essentially demodulation. Once waveform parameters are estimated, it is possible to synthesize a noise-free interference estimate using Equation 22, which may then serve as $\hat{z}(t)$ directly, or used as $d(t)$ in a feedforward canceler if correction for additional effects (e.g., $g(\tau)$) is required. PES is particularly effective against interference from modern communications systems, where the “finite alphabet” property of digital modulations greatly aids in waveform parameter estimation.

When applicable, PES has three compelling advantages: First, INR_d is not directly limited by the received strength of the interference. Second, there is no ingress of astronomy into reference channel; i.e., $f(\tau) = 0$, thereby ameliorating a primary toxicity concern. Third, an external reference signal (i.e., from an auxiliary antenna) is not required. The principal disadvantage of PES is that the technique is sensitive to the details of the waveform, including the stationarity of the waveform parameters, unlike techniques in which the reference signal is obtained from a reference antenna.

In Ellingson et al. (2001), a feedforward canceler using PES is used to mitigate interference from GLONASS from Australia Telescope Compact Array observations of a spectral line at 1612.15 MHz. Despite $\text{INR}_x \ll 1$, IRR in the range 20 dB to 25 dB is achieved. Other studies involving similar PES-type cancelers include Roshi (2002) for analog (NTSC) broadcast television; Ellingson & Hampson (2003), for L-band air surveillance radar; Lee (2008), addressing a wide variety of analog and digital interference waveforms; Nigra et al. (2010) for the US Global Positioning System (GPS); and Ellingson (2020) for VHF-band US weather radio.

9. CONCLUSIONS

The studies cited in the previous section reach essentially the same top-level conclusion: CTC shows promise, but work is incomplete and there are a variety of problems remaining to be solved. These problems fall in two broad categories: (1) Algorithm design (What is the appropriate algorithm, and how to anticipate levels of performance); and (2) Implementation issues remaining to be understood, quantified, and solved. This paper is an attempt to gain a comprehensive understanding of the first category of problems, and has identified some key elements in the second category of problems.

We have identified feedforward MMSE, including the reduced complexity version of Section 5, as a good starting point for development of an operational CTC capability for radio astronomy, and we have demonstrated that this strategy can plausibly meet the requirements for the “look through” capability envisioned in Sections 1 and 2. Along the way we have defined the relevant and useful performance metrics $\overline{\text{IRR}}_1$, $\overline{\text{IRR}}_2$, and NIR. Finally, we have confirmed and quantified the importance of high

INR_d for effective CTC, and identified several strategies by which this can be achieved even in scenarios where INR_x is low.

ACKNOWLEDGMENTS

This paper is based upon work supported in part by the National Science Foundation under Grant ECCS-2029948.

APPENDIX

A. INTERFERENCE REJECTION RATIO

In this appendix we present expressions for the interference rejection ratios $\overline{\text{IRR}}_1$ and $\overline{\text{IRR}}_2$, defined in Section 3.2. In Section A.1, these expressions are derived for feedforward MMSE for the special case of a length-1 filter ($M = 1$) and a single narrowband interferer. In Section A.2, empirical expressions are proposed for the $M > 1$ case. In Section A.3, expressions are derived for LMS with $M = 1$ and a single narrowband interferer.

A.1. Feedforward MMSE, $M = 1$

For notational convenience let us define $x[k] = x(kT_S)$, $z[k] = z(kT_S)$, and so on. As in Section 3, we assume $s(kT_S)$ is negligible in this analysis; i.e.,

$$x[k] = z[k] + n[k] \quad (\text{A1})$$

For convenience and without loss of generality, $z[k]$ is assumed to have unit time-average power and $n[k]$ is assumed to be complex white Gaussian noise (WGN) with variance $\sigma_n^2 = 1/\text{INR}_x$. We further assume $g(\tau)$ is a complex-valued constant with phase θ such that

$$d[k] = \sqrt{\text{INR}_d} e^{j\theta} z[k] + u[k] \quad (\text{A2})$$

where $j = \sqrt{-1}$ and $u[k]$ is unit power complex WGN. In feedforward MMSE, we have

$$\hat{z}[k] = \mathbf{w}^H \mathbf{d}[k] \quad (\text{A3})$$

where \mathbf{w} is the solution to

$$\mathbf{R}\mathbf{w} = \mathbf{r} \quad (\text{A4})$$

In the context of stochastic analysis, time averages are more appropriately expressed as expectations over k . Thus, Equations 11 and 12 of Section 3 become

$$\mathbf{R} = E\{\mathbf{d}[k]\mathbf{d}^H[k]\} \quad (\text{A5})$$

$$\mathbf{r} = E\{x^*[k]\mathbf{d}[k]\} \quad (\text{A6})$$

respectively.

In the special case of $M = 1$ and asymptotically large L , we have:

$$\mathbf{R} = E\{d[k]d^*[k]\} = \text{INR}_d + 1 \quad (\text{A7})$$

$$\mathbf{r} = E\{x^*[k]d[k]\} = E\{(z^*[k] + n^*[k])(\sqrt{\text{INR}_d} e^{j\theta} z[k] + u[k])\} = \sqrt{\text{INR}_d} e^{j\theta} \quad (\text{A8})$$

which, being 1×1 , we shall henceforth refer to simply as “ R ” and “ r ”, respectively. Subsequently,

$$\mathbf{w} = \frac{\sqrt{\text{INR}_d} e^{j\theta}}{\text{INR}_d + 1} \quad (\text{A9})$$

which, also being 1×1 , we henceforth refer to simply as “ w ”. Using these findings in Equation A3, we obtain:

$$\hat{z}[k] = w^* d[k] \quad (\text{A10})$$

$$= \frac{\sqrt{\text{INR}_d} e^{-j\theta}}{\text{INR}_d + 1} \left(\sqrt{\text{INR}_d} e^{j\theta} z[k] + u[k] \right) \quad (\text{A11})$$

$$= \frac{\text{INR}_d}{\text{INR}_d + 1} z[k] + \frac{\sqrt{\text{INR}_d}}{\text{INR}_d + 1} \tilde{u}[k] \quad (\text{A12})$$

where $\tilde{u}[k]$ has been defined as $u[k] e^{j\theta}$ for notational convenience.

Section 3.2 describes two possible definitions of interference rejection ratio; namely IRR_1 (Equation 14) and IRR_2 (Equation 15). Let us begin with IRR_2 . In this case we define:

$$\overline{\text{IRR}}_2 = \frac{E\{|z[k]|^2\}}{E\{z[k] - \hat{z}[k]\}^2} \quad (\text{A13})$$

The distinction between IRR_2 and $\overline{\text{IRR}}_2$ is important: IRR_2 is a measurable outcome from a single trial, whereas $\overline{\text{IRR}}_2$ is a statistic determined from all trials. The latter can be defined in multiple ways; we choose Equation A13 because it facilitates the simple derivation below (alternative definitions lead to much more difficult analysis), and also because Equation A13 is not significantly biased by the intermittent spuriously large values of IRR that are encountered in experiments in which the interferer is a deterministic signal with slowly-varying waveform parameters.

Since we earlier specified $z[k]$ to have unit time-average power, the numerator of Equation A13 is 1. In the denominator, we find:

$$E\{|z[k] - \hat{z}[k]|^2\} = E\left\{\left|z[k] - \frac{\text{INR}_d}{\text{INR}_d + 1} z[k] + \frac{\sqrt{\text{INR}_d}}{\text{INR}_d + 1} \tilde{u}[k]\right|^2\right\} \quad (\text{A14})$$

$$= \left(1 - \frac{\text{INR}_d}{\text{INR}_d + 1}\right)^2 + \frac{\text{INR}_d}{(\text{INR}_d + 1)^2} \quad (\text{A15})$$

$$= \frac{1}{(\text{INR}_d + 1)^2} + \frac{\text{INR}_d}{(\text{INR}_d + 1)^2} \quad (\text{A16})$$

$$= \frac{1}{\text{INR}_d + 1} \quad (\text{A17})$$

Therefore

$$\overline{\text{IRR}}_2 = \text{INR}_d + 1 \quad (\text{asymptotically large } L) \quad (\text{A18})$$

As expected, $\overline{\text{IRR}}_2=1$ for $\text{INR}_d = 0$, and $\overline{\text{IRR}}_2 \rightarrow \infty$ for $\text{INR}_d \rightarrow \infty$. Note also that $\overline{\text{IRR}}_2$ under these assumptions is independent of INR_x , since any limitation due to finite INR_x is made irrelevant by the unlimited observation time (L).

Now we wish to account for the fact that R and r must be estimated from a limited number of samples; i.e., potentially small L . To begin, note that the quality of the estimate of r depends on both INR_d and INR_x , whereas the quality of the estimate of R depends only on INR_d . With this in mind, let us assume that INR_d is large enough that performance is limited primarily by the quality of the estimate of r ; i.e., that the quality of estimation of R has negligible effect in comparison. The quantity $r = E\{x^*[k]d[k]\}$ is estimated from L samples as follows:

$$r = \frac{1}{L} \sum_{k=1}^L x^*[k]d[k] \quad (\text{A19})$$

Substituting for $x[k]$ and $d[k]$ we find:

$$r = \frac{1}{L} \sum_{k=1}^L (z^*[k] + n^*[k])(\sqrt{\text{INR}_d} e^{j\theta} z[k] + u[k]) \quad (\text{A20})$$

$$= \frac{1}{L} \sum_{k=1}^L \sqrt{\text{INR}_d} e^{j\theta} |z[k]|^2 + \frac{1}{L} \sum_{k=1}^L n^*[k]u[k] + \frac{1}{L} \sum_{k=1}^L \sqrt{\text{INR}_d} e^{j\theta} z[k]n^*[k] + \frac{1}{L} \sum_{k=1}^L z^*[k]u[k] \quad (\text{A21})$$

Since we previously set the variance of $z[k]$ to one, the first term reduces to $\sqrt{\text{INR}_d} e^{j\theta}$. The second term is negligible since $n[k]$ and $u[k]$ are uncorrelated. The last two terms can be approximated as statistically-independent Gaussian random variables with zero mean and variances $\text{INR}_d/(L \cdot \text{INR}_x)$ and $1/L$, respectively. Therefore, the sum of the last two terms can be approximated as a Gaussian random variable \tilde{v} with zero mean and variance $\text{INR}_d/(L \cdot \text{INR}_x) + 1/L$. Thus, we may interpret r as a random variable:

$$r = \sqrt{\text{INR}_d} e^{j\theta} + \tilde{v} \quad (\text{A22})$$

Subsequently, the revised expression for w as a random variable which accounts for limited number of samples L is

$$w = \frac{r}{R} = \frac{\sqrt{\text{INR}_d} e^{j\theta} + \tilde{v}}{\text{INR}_d + 1} \quad (\text{A23})$$

and the associated expression for the interference estimate is

$$\hat{z}[k] = w^* d[k] = \frac{\sqrt{\text{INR}_d} e^{-j\theta} + \tilde{v}}{\text{INR}_d + 1} \left(\sqrt{\text{INR}_d} e^{j\theta} z[k] + u[k] \right) \quad (\text{A24})$$

and the denominator of $\overline{\text{IRR}}_2$ becomes

$$E\{|z[k] - \hat{z}[k]|^2\} = E \left\{ \left| z[k] - \frac{\sqrt{\text{INR}_d} e^{-j\theta} + \tilde{v}}{\text{INR}_d + 1} \left(\sqrt{\text{INR}_d} e^{j\theta} z[k] + u[k] \right) \right|^2 \right\} \quad (\text{A25})$$

$$= E \left\{ \left| z[k] \left(1 - \frac{\text{INR}_d}{\text{INR}_d + 1} \right) - \frac{\sqrt{\text{INR}_d}}{\text{INR}_d + 1} e^{j\theta} z[k] \tilde{v} \dots \right. \right. \\ \left. \left. - \frac{\sqrt{\text{INR}_d}}{\text{INR}_d + 1} e^{-j\theta} u[k] - \frac{\tilde{v}}{\text{INR}_d + 1} u[k] \right|^2 \right\} \quad (\text{A26})$$

Neglecting terms corresponding to correlations between uncorrelated noise waveforms, we obtain

$$E\{|z[k] - \hat{z}[k]|^2\} = \left(1 - \frac{\text{INR}_d}{\text{INR}_d + 1} \right)^2 + \left(\frac{\text{INR}_d}{\text{INR}_d + 1} \right)^2 \frac{1}{L \cdot \text{INR}_x} + \frac{\text{INR}_d}{L(\text{INR}_d + 1)^2} + \frac{\text{INR}_d}{(\text{INR}_d + 1)^2} \quad (\text{A27})$$

Thus we obtain

$$\overline{\text{IRR}}_2 = \left[\left(1 - \frac{\text{INR}_d}{\text{INR}_d + 1} \right)^2 + \left(\frac{\text{INR}_d}{\text{INR}_d + 1} \right)^2 \frac{1}{L \cdot \text{INR}_x} + \frac{\text{INR}_d}{L(\text{INR}_d + 1)^2} + \frac{\text{INR}_d}{(\text{INR}_d + 1)^2} \right]^{-1} \quad (\text{A28})$$

which simplifies to

$$\overline{\text{IRR}}_2 = \frac{\text{INR}_x L (\text{INR}_d + 1)^2}{\text{INR}_x L (\text{INR}_d + 1) + \text{INR}_d (\text{INR}_d + \text{INR}_x)} \quad (\text{A29})$$

This yields Equation A18 as expected when either $L \rightarrow \infty$ or $\text{INR}_x \rightarrow \infty$. Of particular interest is the result in the high- and low- INR_d regimes. Note:

$$\overline{\text{IRR}}_2 \rightarrow \text{INR}_x L \quad \text{for } \text{INR}_d \gg \text{INR}_x L \quad (\text{A30})$$

$$\overline{\text{IRR}}_2 \rightarrow \text{INR}_d + 1 \quad \text{for } \text{INR}_d \ll \text{INR}_x L \quad (\text{A31})$$

Now we consider the alternative definition $\overline{\text{IRR}}_1$, which for $M = 1$ is

$$\overline{\text{IRR}}_1 = \frac{E\{|z[k]|^2\}}{E\{|z[k] - w * g * z[k]|^2\}} \quad (\text{A32})$$

where g is the complex gain of the interference in the reference channel; i.e., $d[k] = g z[k] + u[k]$. From previous work we see that we may represent this quantity as $g = \sqrt{\text{INR}_d} e^{j\phi}$ where ϕ is an independent random variable analogous to θ . Assuming for the moment that $z[k]$ is narrowband, $z[k]$ may be factored from the expression yielding:

$$\overline{\text{IRR}}_1 = \frac{1}{E\{|1 - w * g|^2\}} \quad (\text{A33})$$

Following the same analysis as before, we obtain:

$$\overline{\text{IRR}}_1 = \frac{\text{INR}_x L (\text{INR}_d + 1)^2}{\text{INR}_x L + \text{INR}_d (\text{INR}_d + \text{INR}_x)} \quad (\text{A34})$$

Like $\overline{\text{IRR}}_2$, this yields Equation A18 as expected when either $L \rightarrow \infty$ or $\text{INR}_x \rightarrow \infty$, and also

$$\overline{\text{IRR}}_1 \rightarrow \text{INR}_x L \quad \text{for } \text{INR}_d \gg \sqrt{\text{INR}_x L} \quad (\text{A35})$$

However,

$$\overline{\text{IRR}}_1 \rightarrow (\text{INR}_d + 1)^2 \quad \text{for } \text{INR}_d \ll \sqrt{\text{INR}_x L} \quad (\text{A36})$$

The dramatically larger value of $\overline{\text{IRR}}_1$ relative to $\overline{\text{IRR}}_2$ in the small INR_d regime is due to the fact that $\overline{\text{IRR}}_1$ considers only the change in the interference component of output signal, whereas $\overline{\text{IRR}}_2$ interprets noise injection by the canceler as an additional increase in the interference in the output signal. Approximations made in the above derivation are validated by the agreement with simulation results shown in Section 4.

A.2. Feedforward MMSE, $M > 1$

Derivations for expressions valid for $M > 1$ are not available. Instead we propose the following empirical expressions, which are informed by the $M = 1$ analysis in the previous section. These expressions show excellent agreement with the $M > 1$ simulation data (see e.g. Figures 4 and 5) and reduce as expected to the $M = 1$ expressions.

For $\overline{\text{IRR}}_2$:

$$\overline{\text{IRR}}_2 \approx \frac{\text{INR}_x (L/M) (M \cdot \text{INR}_d + 1)^2}{\text{INR}_x (L/M) (M \cdot \text{INR}_d + 1) + M \cdot \text{INR}_d (M \cdot \text{INR}_d + \text{INR}_x)} \quad (\text{A37})$$

$$\overline{\text{IRR}}_2 \rightarrow \text{INR}_x L / M \quad \text{for } \text{INR}_d \gg \text{INR}_x L \quad (\text{A38})$$

$$\overline{\text{IRR}}_2 \rightarrow M \cdot \text{INR}_d + 1 \quad \text{for } \text{INR}_d \ll \text{INR}_x L / M^2 \quad (\text{A39})$$

For $\overline{\text{IRR}}_1$:

$$\overline{\text{IRR}}_1 \approx \frac{\text{INR}_x L (M \cdot \text{INR}_d + 1)^2}{\text{INR}_x L + M^2 \cdot \text{INR}_d (\text{INR}_d + \text{INR}_x)} \quad (\text{A40})$$

$$\overline{\text{IRR}}_1 \rightarrow \text{INR}_x L \quad \text{for } \text{INR}_d \gg \text{INR}_x L \quad (\text{A41})$$

$$\overline{\text{IRR}}_1 \rightarrow (M \cdot \text{INR}_d + 1)^2 \quad \text{for } \text{INR}_d \ll \text{INR}_x L \text{ and } \text{INR}_d \ll L / M^2 \quad (\text{A42})$$

A.3. LMS, $M = 1$

The difference between the theoretical IRR of $M = 1$ LMS and $M = 1$ feedforward MMSE is due to the jitter in w due to the iterative update controlled by μ . For this analysis, we represent w as the random variable $\sqrt{\text{INR}_d} e^{j\theta} + e$, where e represents the noise in the update after convergence, and is well-modeled as WGN. Widrow et al. (1976) have shown that once the algorithm has converged, the power σ_e^2 of this noise is μMSE_{\min} where MSE_{\min} is the minimum mean square error associated with the ideal noise-free solution $w = r/R$. Furthermore,

$$\text{MSE}_{\min} = E \left\{ |x[k] - w^* d[k]|^2 \right\} = \frac{1}{\text{INR}_d + 1} + \frac{1}{\text{INR}_x} = \frac{\text{INR}_d + 1 + \text{INR}_x}{(\text{INR}_d + 1) \text{INR}_x} \quad (\text{A43})$$

In this case we have for the denominator of $\overline{\text{IRR}}_2$, in lieu of Equation A25,

$$\begin{aligned} E \{ |z[k] - \hat{z}[k]|^2 \} &= E \left\{ \left| z[k] - \left(\frac{\sqrt{\text{INR}_d} e^{-j\theta}}{\text{INR}_d + 1} + e^* \right) \left(\sqrt{\text{INR}_d} e^{j\theta} z[k] + u[k] \right) \right|^2 \right\} \\ &= E \left\{ \left| z[k] - \frac{\text{INR}_d}{\text{INR}_d + 1} z[k] - \frac{\sqrt{\text{INR}_d} e^{-j\theta}}{\text{INR}_d + 1} u[k] - \sqrt{\text{INR}_d} e^{j\theta} z[k] e^* + e^* u[k] \right|^2 \right\} \\ &= \frac{1}{\text{INR}_d + 1} + (\text{INR}_d + 1) \sigma_e^2 \end{aligned} \quad (\text{A44})$$

where we have ignored the term representing the product of uncorrelated noise sources. Next we substitute $\sigma_e^2 = \mu \text{MSE}_{\min}$ with MSE_{\min} coming from Equation A43:

$$\begin{aligned} E\{|z[k] - \hat{z}[k]|^2\} &= \frac{1}{\text{INR}_d + 1} + (\text{INR}_d + 1)\mu \frac{\text{INR}_d + 1 + \text{INR}_x}{(\text{INR}_d + 1)\text{INR}_x} \\ &= \frac{1}{\text{INR}_d + 1} + \mu \frac{\text{INR}_d + 1 + \text{INR}_x}{\text{INR}_x} \end{aligned} \quad (\text{A45})$$

Inserting Equation A45 into the definition of $\overline{\text{IRR}}_2$ we have:

$$\overline{\text{IRR}}_2 = \left(\frac{1}{\text{INR}_d + 1} + \mu \frac{\text{INR}_d + 1 + \text{INR}_x}{\text{INR}_x} \right)^{-1} \quad (\text{A46})$$

In the large- and small- INR_d regimes, we find

$$\overline{\text{IRR}}_2 \rightarrow \frac{1}{\mu} \frac{\text{INR}_x}{\text{INR}_d} \quad \text{for } \text{INR}_d \gg \sqrt{\text{INR}_x/\mu} \quad (\text{A47})$$

$$\overline{\text{IRR}}_2 \rightarrow \text{INR}_d + 1 \quad \text{for } \text{INR}_d \ll \sqrt{\text{INR}_x/\mu} \quad (\text{A48})$$

Now we consider the alternative definition $\overline{\text{IRR}}_1$. The analysis for $M = 1$ is the same as in Section A.1 up to Equation A33. Following the same procedure, we find in the case of LMS:

$$\overline{\text{IRR}}_1 = \left(\frac{1}{(\text{INR}_d + 1)^2} + \mu \text{INR}_d \frac{\text{INR}_d + 1 + \text{INR}_x}{(\text{INR}_d + 1)\text{INR}_x} \right)^{-1} \quad (\text{A49})$$

In the large- and small- INR_d regimes, we find

$$\overline{\text{IRR}}_1 \rightarrow \frac{1}{\mu} \frac{\text{INR}_x}{\text{INR}_d} \quad \text{for } \text{INR}_d \gg (\text{INR}_x/\mu)^{1/3} \quad (\text{A50})$$

$$\overline{\text{IRR}}_1 \rightarrow (\text{INR}_d + 1)^2 \quad \text{for } \text{INR}_d \ll (\text{INR}_x/\mu)^{1/3} \quad (\text{A51})$$

We have verified these expressions using simulations under the same conditions as our feedforward MMSE experiments, and have found similarly excellent agreement.

B. DEMONSTRATION USING REAL-WORLD DATA

In this appendix, we demonstrate feedforward MMSE CTC for the mitigation of interference from a terrestrial radio broadcast signal. In this demonstration, we consider the weather radio service of the U.S. National Oceanic and Atmospheric Administration (NOAA). This service is provided by broadcast stations transmitting in 25 kHz channels with center frequencies 162.400 MHz through 162.550 MHz, as shown in Figure 9. Each signal is analog narrowband frequency-modulated voice. Each signal is continuously present, and the instantaneous occupied bandwidth varies dynamically between nearly zero (effectively, a sinusoid) to most of the channel on millisecond timescales. These signals are representative of a great number of sources of terrestrial interference throughout the HF, VHF, and UHF wavebands.

Data was collected from the vicinity of Blacksburg, Virginia, USA using half-wavelength dipoles horizontal to the ground and separated by about 5 m (about 2.7 wavelengths). The signal from each dipole was converted to baseband and sampled at 2.4 million samples per second (MSPS) with 8 bits for “I” and 8 bits for “Q” using a software defined radio with coherent channels. One dipole was aligned in azimuth so as to maximize the 162.450 MHz signal, resulting in the spectrum shown in Figure 9. This signal was filtered (as described below) and served as the reference channel. The other dipole was aligned in azimuth so as to *minimize* the 162.450 MHz signal. This served as the primary channel input, simulating the signal received through a far sidelobe of a radio telescope.⁹ The sensitivity of the receivers is dominated by internal noise, so the noise in the primary and reference channels is uncorrelated.

⁹ The typical level for the far sidelobes of a large reflector is less than 0 dBi; see e.g., ITU Radiocommunication Bureau (2013b).

	$L = 31$	Δ	$L = 104$	Δ	$L = 1042$
$\text{INR}_d/\text{INR}_x L$	+4.45 dB		−0.81 dB		−10.82 dB
IRR_2 (Eq. A29)	+21.52 dB		+24.68 dB		+26.98 dB
IRR (Obs.), 11.52 s	+17.32 dB	−4.20 dB	+22.01 dB	−2.67 dB	> +25.4 dB
IRR (Obs.), 5.76 s	+17.76 dB	−3.76 dB	+23.13 dB	−1.55 dB	> +25.4 dB
IRR (Obs.), 2.88 s	+17.91 dB	−3.61 dB	+23.24 dB	−1.44 dB	> +25.4 dB
IRR (Obs.), 1.44 s	+18.44 dB	−3.08 dB	+24.62 dB	−0.06 dB	> +25.4 dB

Table 4. Summary of predicted and observed CTC performance. “ Δ ” is the ratio of observed performance to the predicted value from Equation A29.

Raw samples were recorded and all subsequent processing was done off-line. First, the 162.450 MHz channel was extracted from the primary and reference channel inputs using a Hamming filter of length 2048 having total bandwidth of 25 kHz. The signals were not downsampled. The resulting primary channel input is shown in the leftmost panel of Figure 10; note that the orientation of the dipoles has resulted in the primary channel INR (INR_x) being much weaker than the reference channel INR (INR_d), as would normally be the case in an operational CTC system.

In order to estimate INR, we estimated noise baselines for the primary and reference channels by log-linear fitting to the noise in the unoccupied 162.3125–162.3875 MHz and 162.5625–162.6375 MHz regions of the spectrum. Using the extrapolated noise baseline to estimate noise power N in the channel, interference power may then be estimated as the difference between total power in the channel $I + N$ and N . Using this method, we estimate $\text{INR}_x = +7.96$ dB and $\text{INR}_d = +27.32$ dB within the 25 kHz channel of interest.

The primary channel is processed using $M = 1$ feedforward MMSE. A single training period is used, but the length of the training period is varied. Training period lengths of 3×10^3 , 1×10^4 , and 1×10^5 samples are considered, corresponding to 1.25 ms, 4.17 ms, and 41.7 ms, respectively. Because the signals were not downsampled after filtering, the corresponding values of L are smaller by the factor (25 kHz)/(2.4 MSPS); i.e., $L = 31$, 104, and 1042; respectively. The estimation filter is calculated once and held constant for the entire duration of the experiment.

The resulting spectra are shown in Figure 10. Clearly CTC is highly effective in this scenario; we see in fact that $L = 1042$ is sufficient to render the interference essentially undetectable. Also note that the noise is neither noticeably increased or noticeably modified. Thus, NIR is too small to be reliably measured, which is consistent with the findings of Section 4.

Now we consider IRR relative to predictions using the theory presented in Appendix A. IRR_2 is the relevant metric for this experiment since only the total power can be measured directly, and the interference power must be estimated using the extrapolated noise baseline as described earlier. Table 4 summarizes the results. First, note that the three values of L considered correspond to INR_d which is high, moderate, and low relative to the $\text{INR}_x L$ criterion identified in Section 4 and Appendix A. The second row of Table 4 shows IRR_2 calculated using Equation A29, which is valid in all three cases. The remaining rows show IRR calculated from total power measurements as described previously. This works sufficiently well for $L = 31$ and $L = 104$, but fails for $L = 1042$ as the difference between interference+noise and noise alone is too small to reliably measure in the $L = 1042$ case.

In order to assess the effect of any non-stationarity of the difference channel ($g(\tau)$ in Equation 3), the experiment is performed with varying dataset lengths ranging from 11.52 s (the entire dataset) down to 1.44 s (the first one-eighth of the dataset), as indicated in the left column of Table 4. When the entire dataset is used, we see that the observed IRR is a few dB less than the theoretical value. However, the observed IRR increases monotonically with decreasing dataset length, suggesting that there is a significant time-varying difference between the propagation channel experienced by the two dipoles over these time scales. Therefore, in this scenario, stationarity considerations require retraining on periods less than a few seconds in order to approach the theoretical limit of Equation A29. This is quite reasonable since even the $L = 1042$ case corresponds to only 41.7 ms of training.

Finally, we conducted an experiment to confirm the “look through” capability of CTC in this scenario, and to assess toxicity. This experiment is summarized in Figure 11. The panel labeled “(b)” shows a simulated astrophysical spectral feature that was generated by filtering uncorrelated noise. This spectral feature is added to the original signal, with the result shown in the panel labeled “(a)+(b)”. The rightmost panel shows the result after CTC with $L = 1042$, using the entire dataset. Note that the spectral feature is recovered with no apparent distortion.

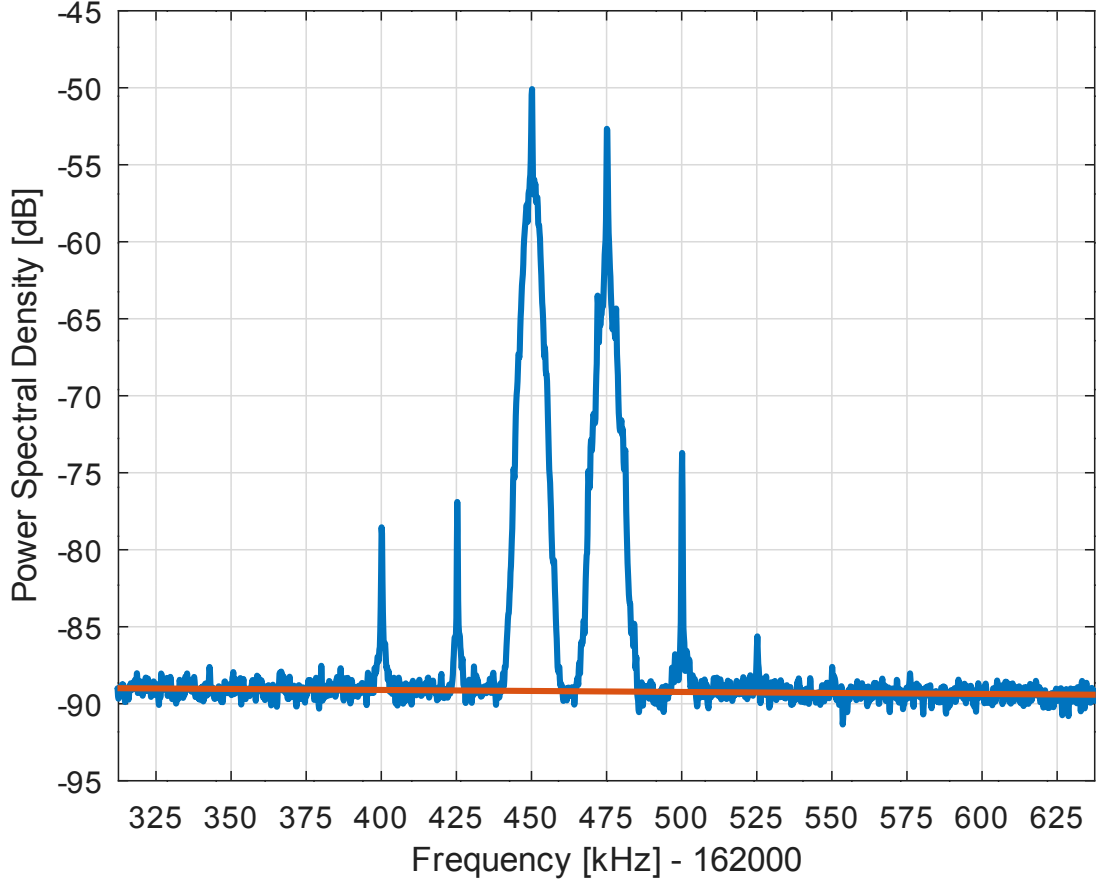


Figure 9. Reference channel ($d(t)$) spectrum prior to channelization. 11.5 s integration, 146.48 Hz spectral resolution, Blackmann-Harris window. The red line is the extrapolated noise floor.

C. EFFECT OF NON-STATIONARITY BETWEEN THE PRIMARY AND REFERENCE CHANNELS

As noted in Section 7.1, the performance of MMSE-based CTC is degraded if the impulse response $g(\tau)$, as defined in Equation 3, is not constant; i.e., non-stationary. In this case we have $g(\tau, t)$; i.e., the function $g(\tau)$ is itself a function of time. When this form of non-stationarity becomes significant, performance will depend on the specific way in which $g(\tau, t)$ is changing with t , and also on what fraction of the time between updates of the filter is used for training. However it is possible to calculate the impact on IRR for $M = 1$ as we shall now show.

For this analysis we assume that the period between filter updates is equal to the L -sample training period. When $M = 1$, $g(\tau, t)$ reduces in the data model to a time-varying complex-valued constant $g_0(t)$. Assuming the same normalization of signals as in Section A.1, the mean square variation due to non-stationarity will be

$$\epsilon^2 = E \left\{ |g_0(t) - \overline{g_0}|^2 \right\} \quad (\text{C52})$$

where the expectation is taken over L samples, and $\overline{g_0}$ is the mean of $g_0(t)$ over these samples. Considering first IRR_2 , Equation A27 becomes

$$E\{|z[k] - \hat{z}[k]|^2\} = \left(1 - \frac{\text{INR}_d}{\text{INR}_d + 1}\right)^2 + \left(\frac{\text{INR}_d}{\text{INR}_d + 1}\right)^2 \frac{1}{L \cdot \text{INR}_x} + \frac{\text{INR}_d}{L(\text{INR}_d + 1)^2} + \frac{\text{INR}_d}{(\text{INR}_d + 1)^2} + \epsilon^2 \quad (\text{C53})$$

Subsequently Equation A29 becomes:

$$\overline{\text{IRR}}_2 = \frac{\text{INR}_x L (\text{INR}_d + 1)^2}{\text{INR}_x L (\text{INR}_d + 1) + \text{INR}_d (\text{INR}_d + \text{INR}_x) + \epsilon^2 \text{INR}_x L (\text{INR}_d + 1)^2} \quad (\text{C54})$$

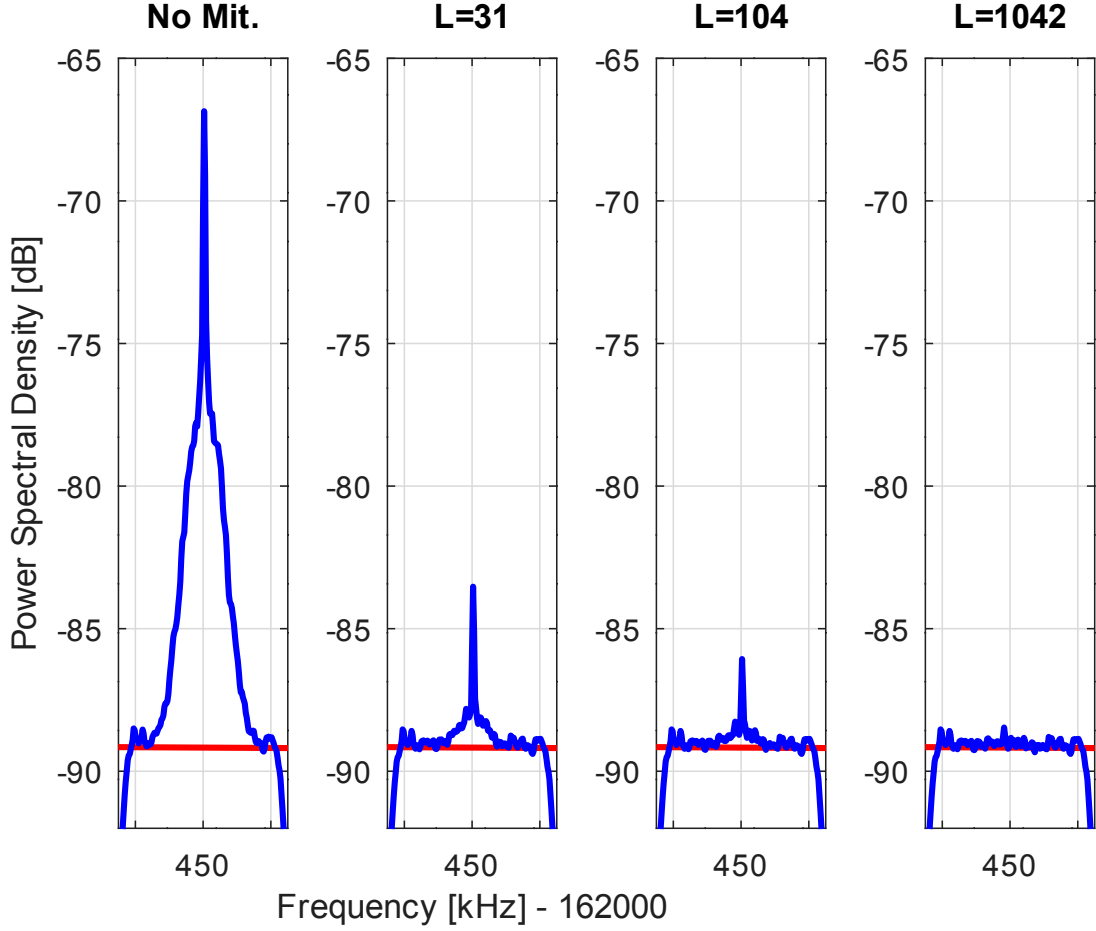


Figure 10. Processing of the 162.450 MHz signal. “No Mit.” is the primary channel ($x(t)$), and the remaining panels show the output ($y(t)$) after feedforward MMSE CTC with the indicated number of training samples (corrected to account for oversampling as described in the text). All other parameters as in Figure 9.

Similarly, for IRR_1 , Equation A34 becomes

$$\overline{\text{IRR}}_1 = \frac{\text{INR}_x L (\text{INR}_d + 1)^2}{\text{INR}_x L + \text{INR}_d (\text{INR}_d + \text{INR}_x) + \epsilon^2 \text{INR}_x L (\text{INR}_d + 1)^2} \quad (\text{C55})$$

We have tested these expressions against simulations in which $g_0(t)$ varies linearly in magnitude with t , and have found excellent agreement.

Finally, let us consider how bad the non-stationarity must be to have a significant effect on IRR. For $\overline{\text{IRR}}_2$, the effect of the term containing ϵ in the denominator of Equation C54 is negligible if

$$\epsilon^2 \ll \frac{\text{INR}_x L (\text{INR}_d + 1) + \text{INR}_d (\text{INR}_d + \text{INR}_x)}{\text{INR}_x L (\text{INR}_d + 1)^2} \quad (\text{C56})$$

The right side of this inequality is simply $1/\overline{\text{IRR}}_2$ evaluated for $\epsilon = 0$. Therefore the effect of non-stationarity is negligible if

$$\epsilon^2 \ll (\overline{\text{IRR}}_2|_{\epsilon=0})^{-1} \quad (\text{C57})$$

The same result is obtained for $\overline{\text{IRR}}_1$, so we may say generally that the effect of non-stationarity is negligible if

$$\epsilon^2 \ll (\text{IRR}|_{\epsilon=0})^{-1} \quad (\text{C58})$$

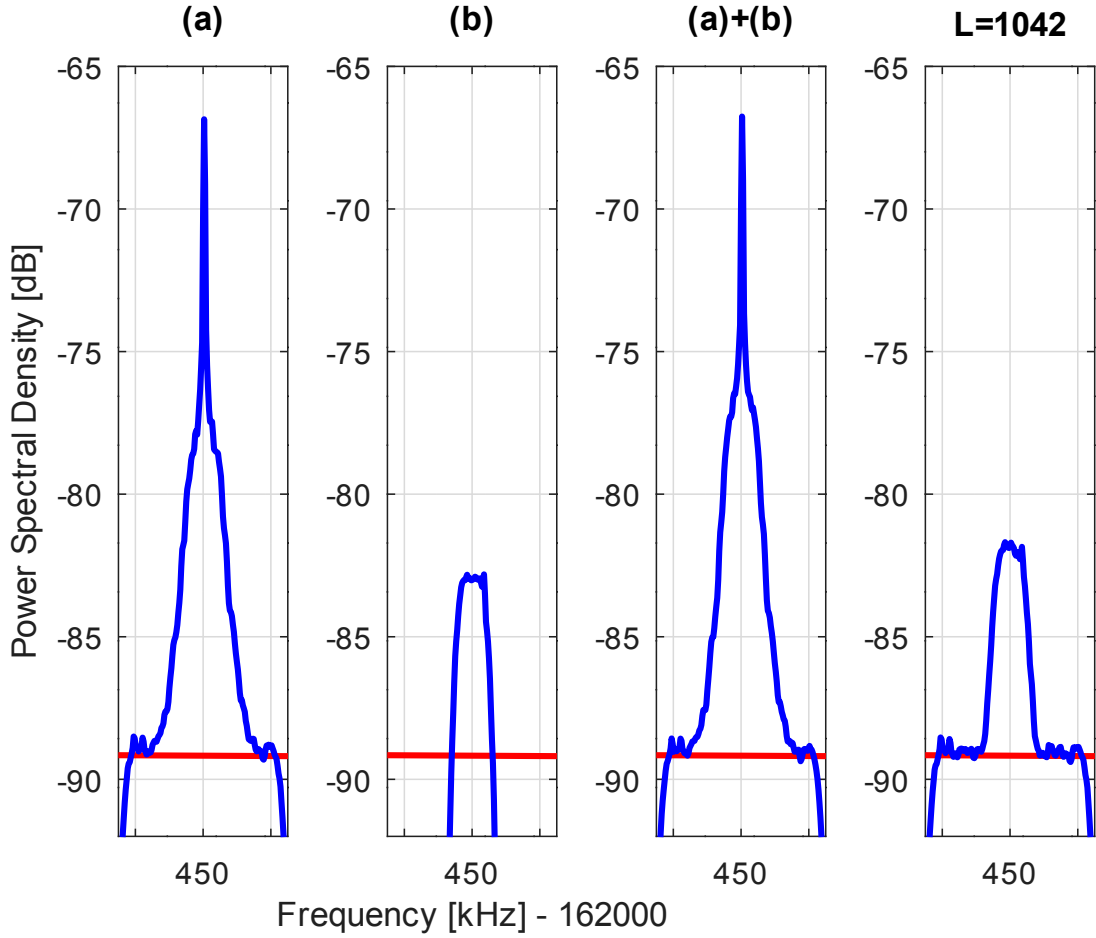


Figure 11. Processing of the 162.450 MHz signal, now with a simulated spectral feature added to the primary channel. (a) and (b) show the original signal and the simulated spectral feature, separately. “(a)+(b)” is the primary channel including the simulated spectral feature. “ $L = 1042$ ” is the output after feedforward MMSE CTC with $L = 1042$. All other parameters as in Figure 9 and 10.

Summarizing, the impact of non-stationarity is greatest when IRR is high, decreases with decreasing IRR, and is negligible when Inequality C58 is satisfied.

REFERENCES

- Barnbaum, C., & Bradley, R. F. 1998, *The Astronomical Journal*, 116, 2598, doi: [10.1086/300604](https://doi.org/10.1086/300604)
- DeGroat, R., Dowling, E., & Linebarger, D. A. 2010, in *The Digital Signal Processing Handbook*, 2nd edn., ed. V. Madisetti (Boca Raton: CRC)
- Ellingson, S. 2002, in *IEEE International Geoscience and Remote Sensing Symposium*, Vol. 3 (IEEE), 1685–1687, doi: [10.1109/igarss.2002.1026221](https://doi.org/10.1109/igarss.2002.1026221)
- Ellingson, S. 2020, *Characterization and Mitigation of Sinusoidal Carrier Modulation Signals by Short-Term Sinusoidal Analysis*. <https://arxiv.org/abs/2004.03649>
- Ellingson, S., & Hampson, G. 2003, *The Astrophysical Journal Supplement Series*, 147, 167, doi: [10.1086/375025](https://doi.org/10.1086/375025)
- Ellingson, S. W., Bunton, J. D., & Bell, J. F. 2001, *The Astrophysical Journal Supplement Series*, 135, 87, doi: [10.1086/321780](https://doi.org/10.1086/321780)
- European Conference of Postal and Telecommunications Administrations (CEPT). 2019, *Compatibility and sharing studies related to NGSO satellite systems operating in the FSS bands 10.7-12.75 GHz (space-to-Earth) and 14-14.5 GHz (Earth-to-space) (ECC Report 271)*
- Ghose, R. 1996, *Interference Mitigation: Theory and Application* (IEEE Press)
- Haykin, S. 2001, *Adaptive Filter Theory*, 4th edn. (Prentice-Hall)

- ITU Radiocommunication Bureau. 2003, Protection criteria used for radio astronomical measurements (ITU RA.769), 2nd edn. (International Telecommunications Union)
- . 2013a, Techniques for mitigation of radio frequency interference in radio astronomy (ITU RA.2126-1), 2nd edn. (International Telecommunications Union).
<https://www.itu.int/pub/R-REP-RA.2126-1-2013>
- . 2013b, Space research earth station and radio astronomy reference antenna radiation pattern for use in interference calculations, including coordination procedures, for frequencies less than 30 GHz (ITU-R SA.509-3), 3rd edn. (International Telecommunications Union).
<https://www.itu.int/rec/R-REC-SA.509-3-201312-I>
- . 2015, Levels of data loss to radio astronomy observations and percentage-of-time criteria resulting from degradation by interference for frequency bands allocated to the radio astronomy service on a primary basis (ITU RA.1513), 2nd edn. (International Telecommunications Union).
<https://www.itu.int/rec/R-REC-RA.1513>
- Kesteven, M., Hobbs, G., Clement, R., et al. 2005, Radio Science, 40, doi: <https://doi.org/10.1029/2004RS003136>
- Kodheli, O., Lagunas, E., Maturo, N., et al. 2020, Satellite Communications in the New Space Era: A Survey and Future Challenges. <https://arxiv.org/abs/2002.08811>
- Lee, K. 2008, PhD thesis, Virginia Tech.
<http://hdl.handle.net/10919/28929>
- National Research Council. 2010, Spectrum Management for Science in the 21st Century (Washington, DC: The National Academies Press), doi: [10.17226/12800](https://doi.org/10.17226/12800)
- Nigra, L., Lewis, M., Edgar, C., et al. 2010, in Proceedings of RFI mitigation workshop — PoS(RFI2010) (Sissa Medialab), doi: [10.22323/1.107.0025](https://doi.org/10.22323/1.107.0025)
- Offringa, A., van de Gronde, J., & Roerdink, J. 2012, A&A, 539, A95, doi: [10.1051/0004-6361/201118497](https://doi.org/10.1051/0004-6361/201118497)
- Poulsen, A. J. 2003, Master's thesis, Brigham Young University. <https://scholarsarchive.byu.edu/etd/92/>
- Roshi, D. 2002, Cancellation of TV Interference, Tech. rep., NRAO
- United Nations Office of Outer Space Affairs and the International Astronomical Union. 2021, Dark and Quiet Skies for Science and Society: Report and Recommendations.
<https://noirlab.edu/public/products/techdocs/techdoc021/>
- Widrow, B., McCool, J., Larimore, M., & Johnson, C. 1976, Proceedings of the IEEE, 64, 1151, doi: [10.1109/PROC.1976.10286](https://doi.org/10.1109/PROC.1976.10286)
- Yang, B. 1995, IEEE Transactions on Signal Processing, 43, 95, doi: [10.1109/78.365290](https://doi.org/10.1109/78.365290)