Part I – Introduction

1 Preface

This book describes a few [sub]optimal adaptive filtration algorithms for solving various room acoustic related audio / speech processing problems such as Adaptive Echo / Noise / Feedback / Reverberation Cancellation, etc. If you are reading this text, you most likely already know what AEC / ANC / AFC are about. Other applications may exist but I am not an expert in other fields.

Sub-optimal algorithms are a can of worms, you can start working on them but you can never finish. Thus, there is nothing conclusive nor final in this publication, it is a set of intermediate results for the next researcher(s) to pick up and evolve further, and I would like to pre-emptively apologize for the lengthy explanation in the "*We shall tell it at length, thoroughly, in detail – for when did a narrative seem too long or too short by reason of actual time of space it took up? We do not fear being called meticulous, inclining as we do to the view that only the exhaustive can be truly interesting*" style instead of "*For every complex problem there is an answer that is clear, simple, and wrong*".

The main problem addressed here is the curse of dimensionality and close-to-singular spectra, in the context of real-time low latency slightly nonlinear, both stationary and non-stationary FIR system identification of Room Impulse Response (RIR). The base is the well-developed theory of adaptive control.

The proposed Fast(er) Subband Adaptive Filtering (FSAF) is an evolution of the Subband Adaptive Filtering (SAF) approach originally proposed by Prof. Dr.-Ing. Walter Kellermann in "Analysis and design of multirate systems for the cancellation of acoustical echoes" at ICASSP-1988. Although the text below is full of comparisons with the original proposal by Dr. W. Kellerman and demonstrations of the superiority of the new approach, these comparisons shall not be considered as belittling of the original approach in any respect.

The set of new techniques, summarily named FSAF, is faster in all respects: converging faster, taking less MIPS, having lower processing latency, etc. But it's NOT as "fast" as FFT where all opportunities have been exploited. The proposed technology is not THE fastest adaptive technology but a step towards it, one of infinitely many. FSAF, besides usual per-subband processing, can be used to solve very high dimension system identification problems in a divide and conquer style. For any predefined precision δ, FSAF solves M much smaller, ~$1/(M + o(1/\delta^p))$, better-conditioned problems in subbands, using either RLS or diagonal / scalar step-size algorithms like Kaczmarz a.k.a. [N]LMS, and converts them back to the full-band time domain, which is Perfect Reconstruction Open-Loop Delay-less SAF. Note that FSAF allows nesting / recomposing of the subband architecture, thus allowing efficient fast converging low-latency low MIPS application for A#C.

In 2001, yours truly left his job and went home, to work on the theoretical background of subband adaptive filtering. It was obvious that all low-hanging fruits have already been harvested, and I needed years of concentration on research to achieve anything of value. Alas, there was not a single employer around willing to wait an undetermined number of years for unpredictable results. Most of the groundwork was done in the early 2000s while working on AEC. By 2004...2005, good understanding of the core problems crystallized. It looked so obvious that I could not believe it had not been found by somebody else… but I could not find a relevant publication. 15 years later, I am still puzzled. In 2019, I decided to "pass the torch" to somebody else[1].

[1] ...due to quite slim chances to recover from a disability. After 2005, I had to start taking contracts and short employments to pay bills. Those contracts / employments were linked to licensing the fruits of my research. It was a way too hard because [in reality] the IP business is only for rich people and corporations. I heard a few times: "You don't have any rights because you don't have enough money to protect them" while dealing with many so-called effective managers. At the end of this voyage, I became disabled with severe depression due to licensing- and work-related problems, thanks to those managers. Major depression is a devastating illness. Please do not take offence if I don't answer your emails / calls.

1.1 WHY MATLAB

It's no secret that >95% of creative ideas coming to curious minds are mistakes. It's very easy to make mistakes while operating with a language, either natural, programming or formal mathematical - but very hard to detect, locate and fix them.

When I was young and clever, I've been covering endless pages with long exotic formulas. Now I became much lazier and find Newton's "visual" approach to be much more constructive. Now, I find the bias towards [error-prone] symbolic linear consecutive narratives to describe the reality (which is, by the way, not necessarily describable as such) to be of questionable value. If 99% of our time goes to fixing mistakes then maybe it's worth trying something different. I even think that we need to "see" the problem prior to diving into any formulations.

My brain certainly does its best with images. MATLAB is the best and the cleanest visualizing analytic tool I know of.

1.2 PREREQUISITES

This book is not an introduction to the field of Digital Signal Processing (DSP), nor to Adaptive Signal Processing or Filtering, nor to Multirate Signal Processing, nor to Subband Adaptive Filtering.

Mandatory prerequisites are:

1. Matlab & Signal Processing Toolbox.
2. F. Gustafsson (2000), Adaptive Filtering and Change Detection, Wiley
3. Crochiere, R.E., and Rabiner, L.R. (1983) Multirate Digital Signal Processing. Prentice-Hall, Englewood Cliffs.
4. R.G. Vaughan, N.L. Scott, D.R. White (1991) The theory of bandpass sampling, IEEE Trans. Signal Processing
5. T.I. Laakso, V. Valimaki, M. Karjalainen, U.K. Laine (1996) Splitting the unit delay. IEEE SPM

For a well-written review of subband adaptive filtering, I'd recommend reading papers by Prof. dr. ir. M. Moonen from Katholieke Universiteit Leuven and/or PhD dissertations of his students.

Prof. Lennart Ljung books, articles, Matlab's "System Identification Toolbox" are not required but wholeheartedly recommended as tremendously enlightening.

The FSAF is discussed here on the example of Acoustic Echo Cancellation (AEC) because it's the simplest application of adaptive filtering. Simple is good… sometimes.

2 SOURCES OF ACOUSTIC ECHO

Acoustic echo is formed when the sound emitted by a speakerphone's loudspeaker gets reflected from the walls, ceilings, floor, furniture, people, etc. back to the speakerphone's microphone. Sound pressure level decreases with each reflection. Some surfaces, such as heavy carpet, soft furniture, open half-full bookshelves with varying format books in random order, people, animals and especially acoustic foam and panels, reflect very little but absorb, dissipate or otherwise significantly attenuate acoustic echo.

Surfaces as glass, brick, gypsum board walls, etc. reflect about 95% of the sound back. The reflections, being repeated multiple times, create reverberation effect[2].

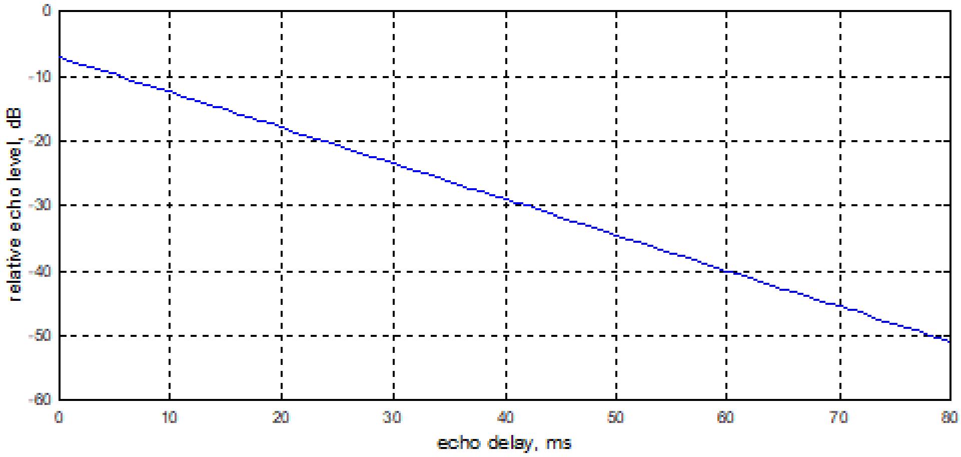
Typically, the reverberation level decreases exponentially with time, so the rooms are often characterized by $RT_{60,}$ which specifies the time when reverberation level drops by 60 dB ($RT_{30} = RT_{60}/2$). For a typical office / conference / living room, RT_{60} lies between 300 and 600 ms, tending to group around 400ms. Shorter RT60 values are preferable but require competent sound room treatment which is rare. Longer RT60 values make conversations inside the conference room itself poorly intelligible, people quickly get fatigued and angry and achieving an agreement becomes an unnecessary hard, exhausting or impossible task. In some poorly designed conference rooms, such as with glass wall-to-wall and floor-to-high-ceiling, RT60 goes up to and above 1000ms.

3 ACOUSTIC ECHO'S EFFECTS ON CONVERSATIONS

The binaural human hearing system copes remarkably well with reverberation effect when both people in conversations are in the same, room[3]. This is not the case if the same people are in different rooms and they use a speakerphone for conversation.

Usually, the echo is most audible on transitions between phonemes and especially to silence, and least audible on long vowels. The perceptibility and annoyance of the echo is influenced by many subjective and objective factors as language spoken, emotional context, importance of particular conversation, hearing skills, etc. The major factor in the delayed echo perceptibility is the delay time between "direct" signal (such as far-end handset's sidetone signal and hearing her/himself) and the "echo", and their relative strengths.

The rough 'rule of thumb' is that the absolute threshold of perceptibility of a delayed single echo of human speech decreases as 0.5 dB per ms of delay (approximately) for the first 40...80ms, then it flattens to -60...65dB, unless the echo is masked by background noise.



With nominal gains in the loudspeaker's and microphone's paths, the acoustic echo is usually about -8…-3 dB in typical rectangular plain-wall office rooms without some soft furniture. The resonating cabinet of

[2] Room acoustic is a complicated issue, but it has been extensively studied in depth from both theoretical and practical perspectives. See specialized literature, as H. Kuttruff, "*Room acoustic*", Applied Science Publishers Ltd, London, 1973, for further and deeper description.

[3] Researchers note that the sound processing by the brain plays a major role in the human hearing system. See: Peter H.Lindsay, Donald A. Norman. "*Human information processing*". University of California, San Diego, Academic Press, NY and London, 1972.

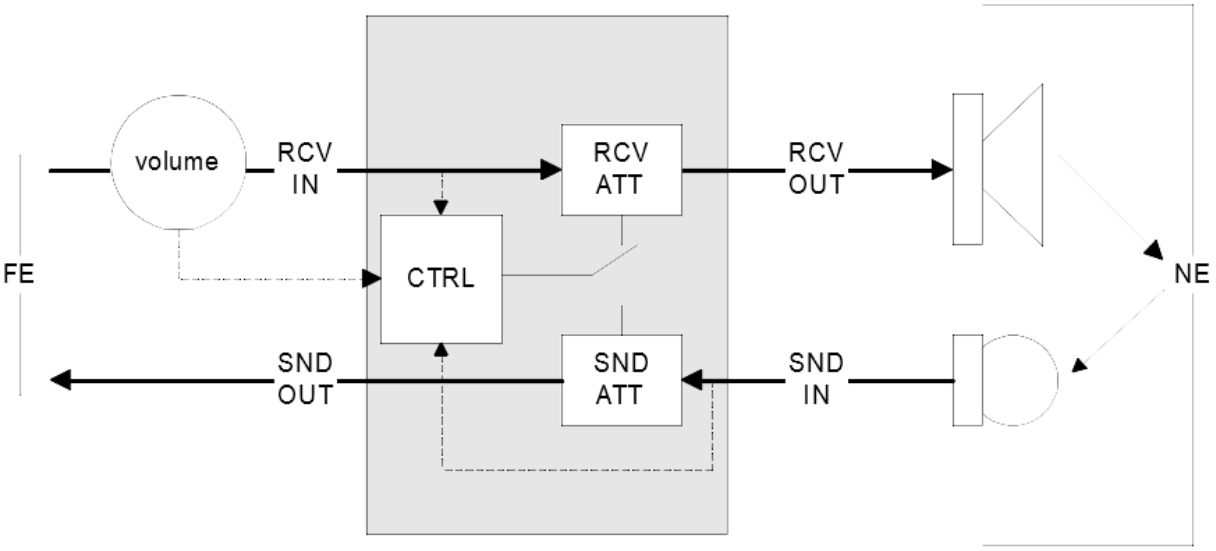
the speakerphone itself usually drives the total echo up to +5…+10 dB. In some conditions, as for tablets and smartphones, mechanical-acoustic echo may be 20+ dB higher than the original received signal.

High-level acoustic echo becomes very annoying and disturbing, and thus it shall be removed to enable handset-like conditions. Moreover, if both people use speakerphones, the acoustic feedback may (and often does) lead to ringing and howling, thus disabling the conversation entirely.

4 Overcoming acoustic echo problems by Acoustic Echo Controller (AEC)

4.1 AEC (controller) Operating Principles

Long before Adaptive Echo Cancellation was introduced, people used Acoustic Echo Controllers, which switched attenuation between microphone and loudspeaker at the Near End (NE). Traditionally, the direction towards the person’s ear is called Receive (RCV), and from the person's mouth – Send (SND).



Basic Acoustic Echo Controller block diagram.

The attenuation, inserted by such basic AEC to make the acoustic echo inaudible, shall account for the maximal strength of the acoustic feedback, RCV path volume setting (amplification or attenuation), and the maximal expected delay on the connections between far-end and near-end.

The cumulative attenuation between RCV IN and SND OUT, a.k.a. Terminal Coupling Loss (TCL), should be high enough (20…30 dB) for low-latency connections and increase as aforementioned 0.5dB/ms of echo delay. Therefore, AEC shall switch at least 45…60 dB of attenuation between SND and RCV.

4.2 DEFICIENCIES

The acoustic echo level is high, thus it is nearly impossible for a basic AEC to distinguish between the echo and true near-end signal. So AEC significantly attenuates the SND path signal whenever RCV is active. When SND activity ends, SND attenuation is gradually switched off to RCV, according to the reverberation time RT_{60} for the target room sizes.

The NE party shall always patiently wait till the far end party ends speaking and the echo dies out – or speak very loudly (nearly shout) into the microphone to enforce AEC to switch SND path attenuation off. Otherwise, the starting unvoiced phoneme may be easily clipped, so “four” may sound as “or”, etc. That is clearly asymmetric and puts the party on the NE to disadvantage in critical conversations, and sometimes leads to conflicts, which could be avoided if the conversations were full duplex.

Note that mild attenuation (less than 10 dB, flat over frequency range) would allow both parties to communicate more or less freely, but any attenuation higher than 25 dB effectively enforces half-duplex operations and automatically gives priority to far-end. In a typical scenario, there is no difference if an AEC inserts 25dB or 55dB of attenuation in the signal path – in either case NE (your) speech intelligibility would be near zero.

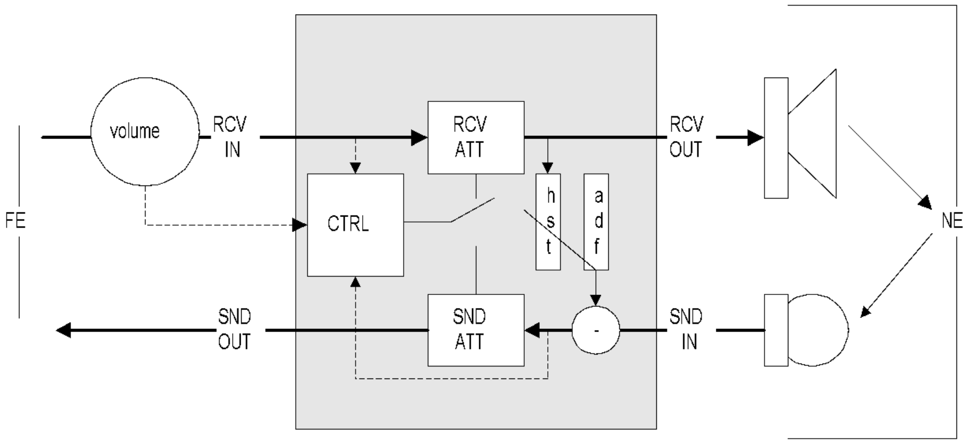
If you desire to be heard, you need to ensure that the echo control algorithm does not attenuate your voice as soon as you start talking. Then, the FE echo control will automatically switch to their RCV (your SND) and your opinion will be heard. Unfortunately, there is no algorithm known that would ensure that your opinion is not only heard but also respected.

5 OVERCOMING DEFICIENCIES OF AEC BY ADDING ECHO CANCELLATION

5.1 AEC (CANCELLER) OPERATING PRINCIPLES

The most obvious way to improve the basic AEC is to add to it an adaptive filter.

Here, we do not attempt to model the room acoustics with meaningful physical models, we use a gray - box Finite Impulse Response (FIR) representation.



ADaptive Filter (ADF) simulates the echo path response in a linear approximation. Adaptive filter is convoluted with the RCV signal history (HST) to obtain the echo estimate, which is then subtracted from SND IN signal. This process is called echo cancellation. The residual error signal is used to tune the adaptive filter with algorithms ranging from NLMS to RLS / Kalman.

If an adaptive filter provides X dB of echo cancellation (an adaptive part of TCL), then the amount of attenuation switched between SND and RCV (a switched part of TCL) can be lowered by the same X amount. In theory, if an adaptive filter can provide echo cancellation equal to the target TCL figure, the attenuation switching becomes unnecessary and an AEC starts to operate in true full-duplex mode.

This direct approach, looking quite sensible, does not lead to significant improvements in practice

5.2 Deficiencies

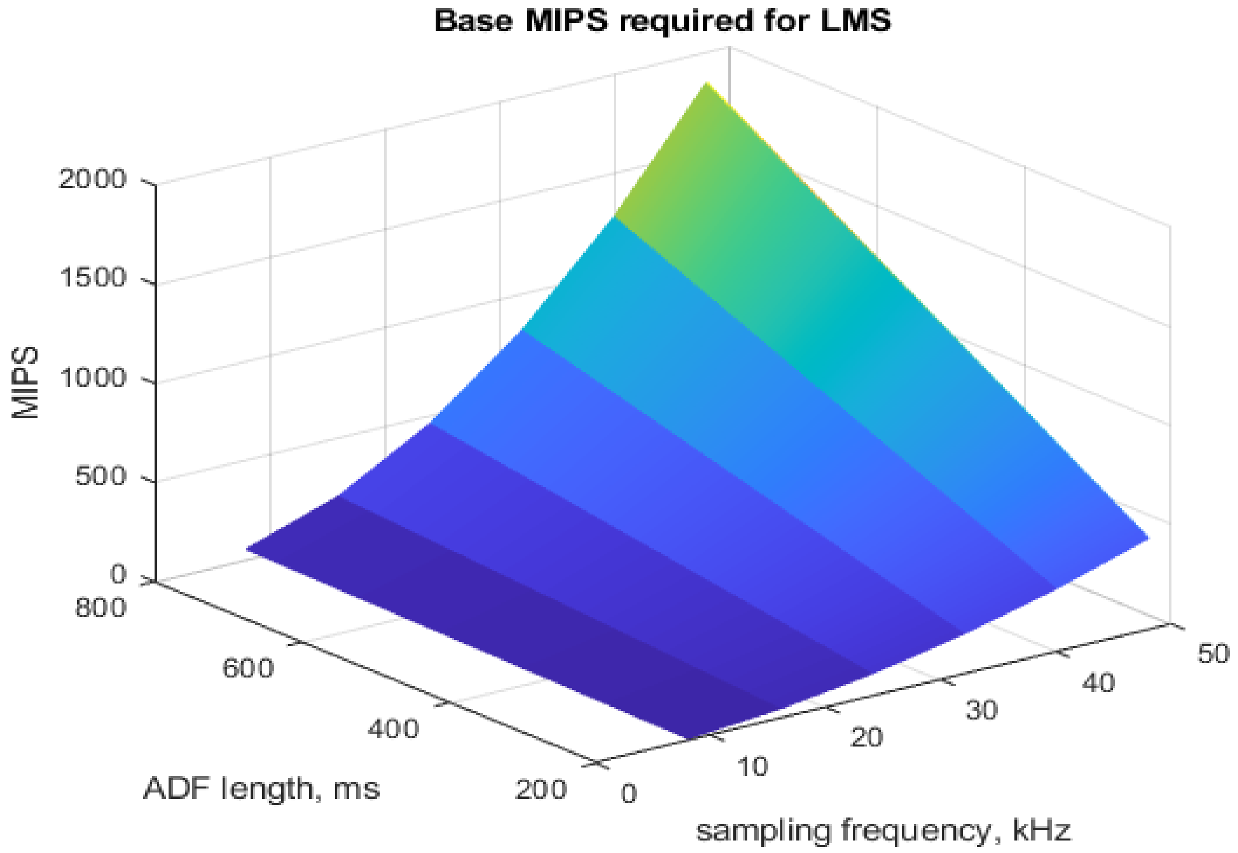
- The adaptive filter needs to be very long to provide 30...40 dB of echo cancellation. Rough estimate for 30dB is at least 0.5* RT_{60}, what corresponds to 150…300 ms. That stipulates very high MIPS load. The longer echo tail is, and the higher is sampling frequency, the more challenges it represents. For LMS, MIPS can be estimated as o(FS*FS*RT60), for RLS, as FS*(FS*RT60^2). For 16kHz sampling we need to model the room with an Adaptive Filter (ADF) of about 8,000 taps if we want to be able to cancel 35+ dB of room echo (the better, the longer), ~130...250 MIPS for LMS and 1,000,000 MIPS for RLS. For more universal 48kHz sampling the model becomes proportionally longer – in order of 20,000+ taps, and requires 1200...2500 MIPS for LMS, and too much for RLS.
- The excitation / input / RCV is usually a human voice or music. Such input is really badly conditioned, which is rather a norm for real-life signals, opposite to what is often used in academic research. Solving a problem with dimension in 10,000s with a nearly singular excitation directly is not a trivial task.
- Second order effects:
  - The speakerphone's acoustic echo path is not stationary and varies with slightest moves of the people in the same room[4] , breathing also can be sensed by a good AEC. The longer AEC's echo tail is, the more pronounced is this effect. We need at least a dual-model approach which doubles MIPS. An IMM with Kalman filter, with its elaborated step size control and Gramm-Schmidt orthogonalization, is indeed required for the AEC, as well as echo path change and double talk detectors.
  - Although the acoustic reflections are quite linear for reasonable sound pressure levels, a typical tablet / laptop / speakerphone's loudspeaker is very far from being a linear transducer[5]. That gives rise to many problems and puts a hard limit on the depth of echo cancellation.
- … and the endless list of third order effects.

[4] C.Antweiler, H.-G.Symanzik, "*Simulation of time variant room impulse responses*", Proceeding of ICASSP' 95, pp. 3031-3034.

[5] R.A. Greiner, T.M.Sims, Jr. "*Loudspeaker distortion reduction*", J. Audio Eng. Soc. Vol. 32, 1984 December, pp. 956-963.

5.2.1 MIPS deficiency details

Let's illustrate the first order effects, starting with the MIPS effect. The figure below is the required MIPS to perform 1 cycle operation with adaptive filter, assuming MAC takes 1 cycle, such as cancellation. Each model of non-delayed LMS would take (load-MAC-save) 3x more, etc [doc_p102.m], and a 2-model LMS-based AEC would take around 10x.



5.2.2 LMS Spectral deficiency details

L = filter length, *h(1:L)* adaptive filter

xs(t) - excitation (spk), *x = xs(t:-1:t-L+1);* let *xs()* contain only 2 orthogonal on length *L* waveforms, complex-domain sines $\mathrm{xs}_1(t)$ and $\mathrm{xs}_2(t)$, mixed in decaying proportions *coef.*

y_t - input (mic), xm(t), $y_t = x^H h$, where h is true impulse response

e(t) - residual error, $e(t) = y_t - x^H h_t$

$$h_{t+1} = h_t + \mu \frac{x}{x^H x}\left(y_t - x^H h_t\right);$$

Then h_t can be represented as a sum of x_1 and x_2, with corresponding projections, and step sizes:

$$\mu_1 = \mu \frac{x_1{}^H x_1}{x_1{}^H x_1 + x_2{}^H x_2} \text{ and } \mu_2 = \mu \frac{x_2{}^H x_2}{x_1{}^H x_1 + x_2{}^H x_2};$$

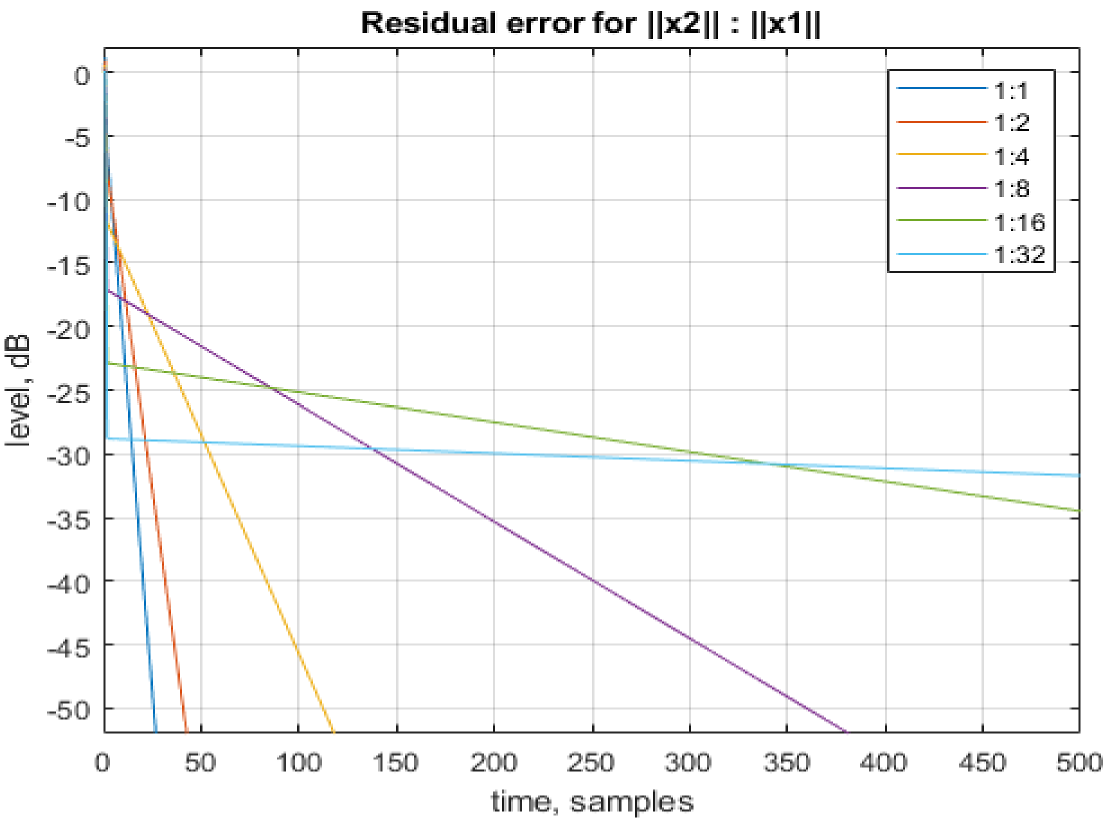
if mixing *coef* is small, the step size for 2nd component will drop as *coef^2;*

$$h_{1,t+1} = h_{1,t} + \mu \frac{x_{1,t}}{{x_1}^H x_1 + {x_2}^H x_2} \left(y_t - \left({x_{1,t}}^H h_{1,t} + {x_{2,t}}^H h_{2,t} \right) \right)$$

$$h_{2,t+1} = h_{2,t} + \mu \frac{x_{2,t}}{{x_1}^H x_1 + {x_2}^H x_2} \left(y_t - \left({x_{1,t}}^H h_{1,t} + {x_{2,t}}^H h_{2,t} \right) \right)$$

So, the MSE for the first components starts high but converges quickly (1step), and MSE for the first components starts low *(coef)* but converges very slowly.

The adaptive filtering theory was initially developed in the context of adaptive control where dimensions are much lower, and the spectrum spread was somewhat a second order effect. In the context of acoustic-related adaptive filtering, it is THE major and central consideration and the famous “curse of dimensionality” reveals itself in all its grandeur [doc_p101.m].



Fisher Matrix eigenvalue spectrum and DFT spectrum are strongly related. Fisher matrix is non-negative symmetric, i.e. an ellipsoid with eigenvectors as axis and eigenvalues as radii, it has the “worst” spread of any orthonormal basises in that space. A sorted eigen spectrum will decay faster and deeper then a sorted DFT spectrum. Thus, if we see a large energy spread in DTF basis, the reality is even worse.

Voiced phonemes’ spectrum is limited to the pitch harmonics, ~100Hz for males and ~250 for females, even higher for children and (a simplification) spreads to 4kHz, so we have 16-25 relatively slow decaying eigenvalues out of ~FS*0.7*RT60, then a sharp drop to near the noise floor. Also consider that an averaged voice spectrum peaks at ~300Hz and then decays 6dB/octave. As the result, simple adding of adaptive filtering to basic AEC may only result in a modest decrease of the amount of attenuation switched between RCV and SND, as 10…15 dB. That does not solve their problems because, from the user’s perspective, there is no perceptible difference between switching 45 dB and 35 dB – both values are too high.

We should also note that LMS is, in a certain sense, memory-less. Let's simplify a bit. After converging to the subspace defined by a certain pitch, it has to reconverge (as from the start) to the new subspace whenever speech with a new pitch comes, even if the RIR stays exactly the same, and the properties of RIR's projection to the first pitch's subspace are lost. If male and female voices alternate, LMS is never quite converged. Although, the human pitch is never constant (as for a flute), and this effect becomes smaller if ADF is long and pitch variability is higher.

As a result, many users find that a simple, but well-tuned acoustic echo controller outperforms adaptive echo cancellers.

6 OVERCOMING DEFICIENCIES OF A BASIC ADAPTIVE AEC BY SAF-1988

6.1 SAF-1988 OPERATING PRINCIPLES

Many improvements over basic adaptive AEC have been proposed in the course of last 30+ years, including frequency[6] (or transform / sub-band / Gabor[7]) domain adaptive filters[8], various post-filtering algorithms[9],[10], ways for exploiting human ear properties[11], etc.

Frequency domain approaches were found slow converging and having high latency, and nothing really worked until Prof. Dr.-Ing. W. Kellermann proposed a novel concept of Subband Adaptive Filtering (SAF) in "Analysis and design of multirate systems for the cancellation of acoustical echoes" on ICASSP-1988.

[6] J/J.Shynk, "*Frequency-domain and multirate adaptive filtering*", IEEE SP Magazine, January 1992, pp. 14-36.

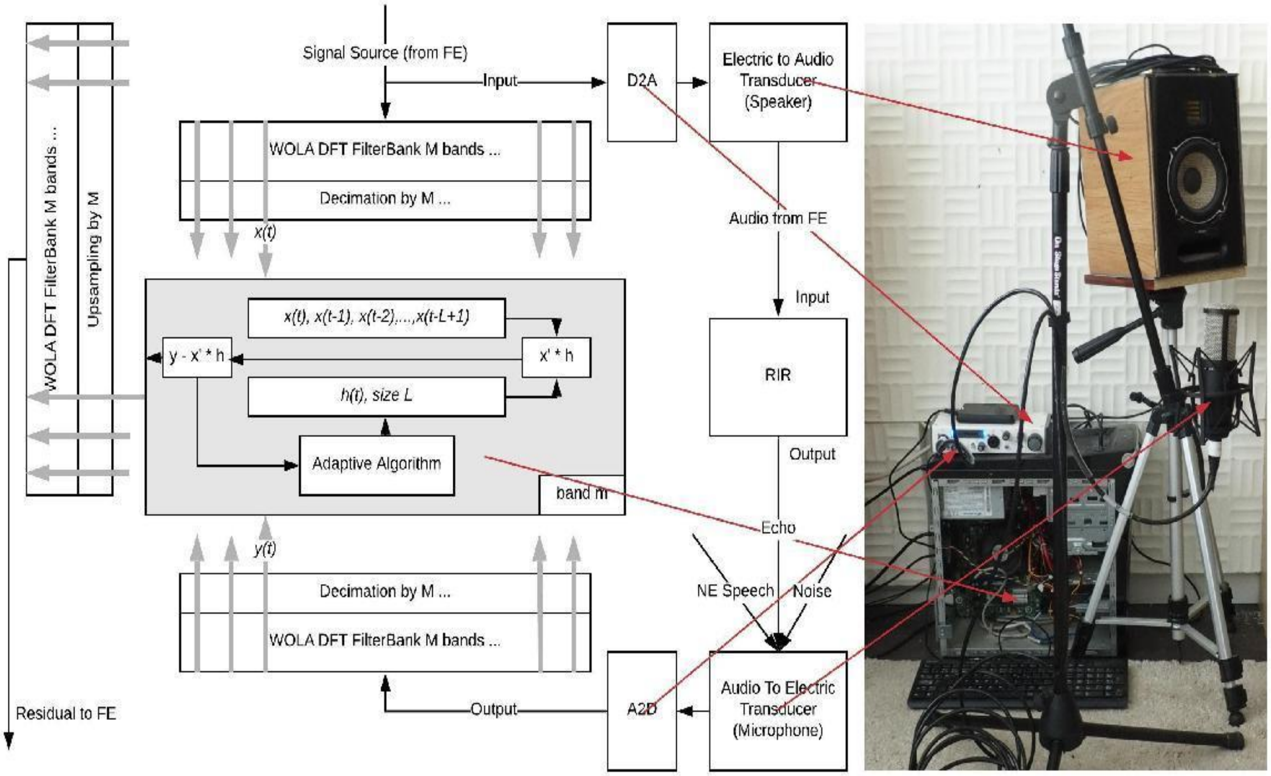
[7] Y.Lu, J.M.Morris, "*Gabor expansions for adaptive echo cancellation*". IEEE Signal Processing Magazine, March 1999, Vol.16, No.2 pp. 68-80.

[8] E. Hansler, G.U.Schmidt. "*Hands-free telephones – joint control of echo cancellation and postfiltering*". Signal Processing 80 (2000) 2295-2305, Elsevier Science B.V.

[9] V.Turbin, A.Gilloire, P. Scalart, "*Comparison of three post-filtering algorithms for residual echo reduction*", Proceedings ICASSP 97, Munich, Germany, 1997, pp.307-310.

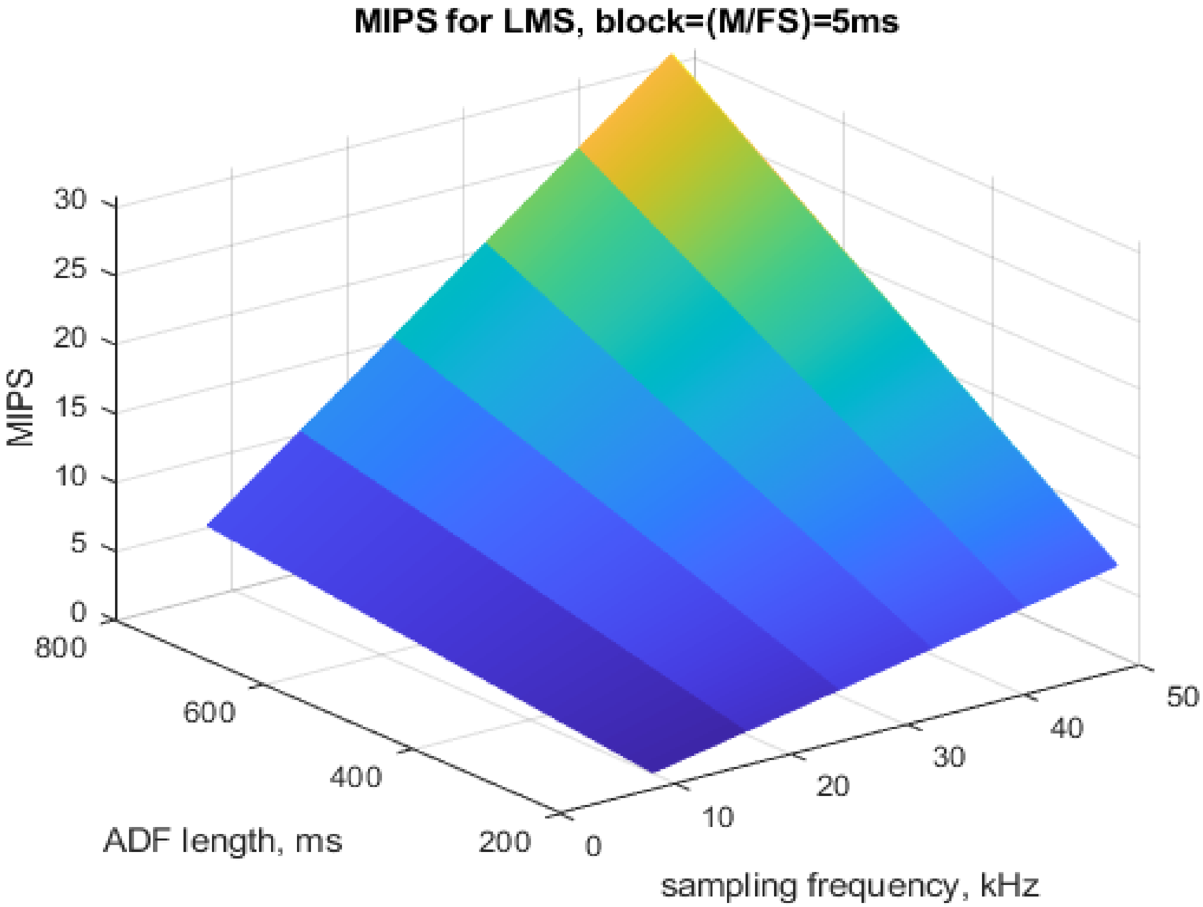
[10] S.Gustafsson, R.Martin, P.Vary, "*Combined acoustic echo control and noise reduction for hands-free telephony*", Signal Processing 27 (3) (1992), pp. 259-271.

[11] V.Turbin, A.Gilloire, P. Scalart, C. Beaugeant, "*Using psychoacoustic criteria in acoustic echo cancellation algorithms*". Proceedings of IWAENC, London, September 1997, pp.53-56.

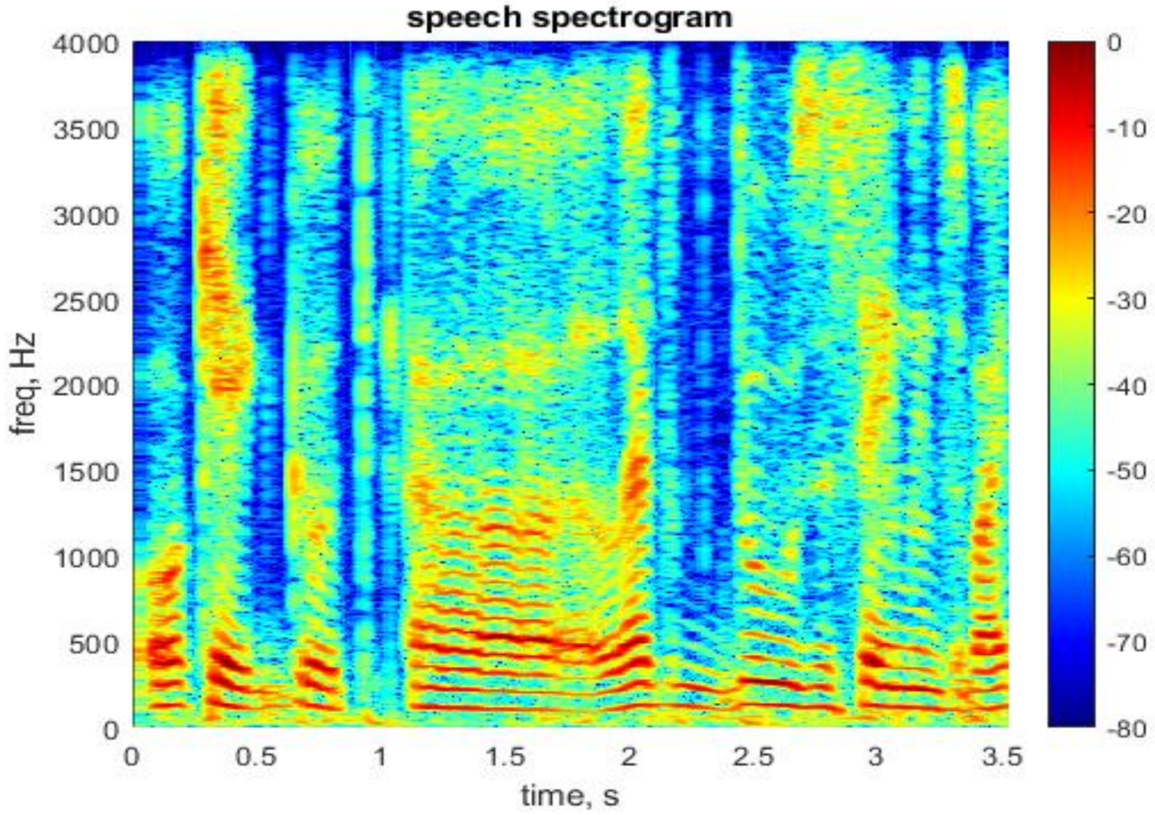


This approach exploited dimension partitioning to successfully address the main shortcomings: eigenvalue spectrum spread and MIPS reduction.

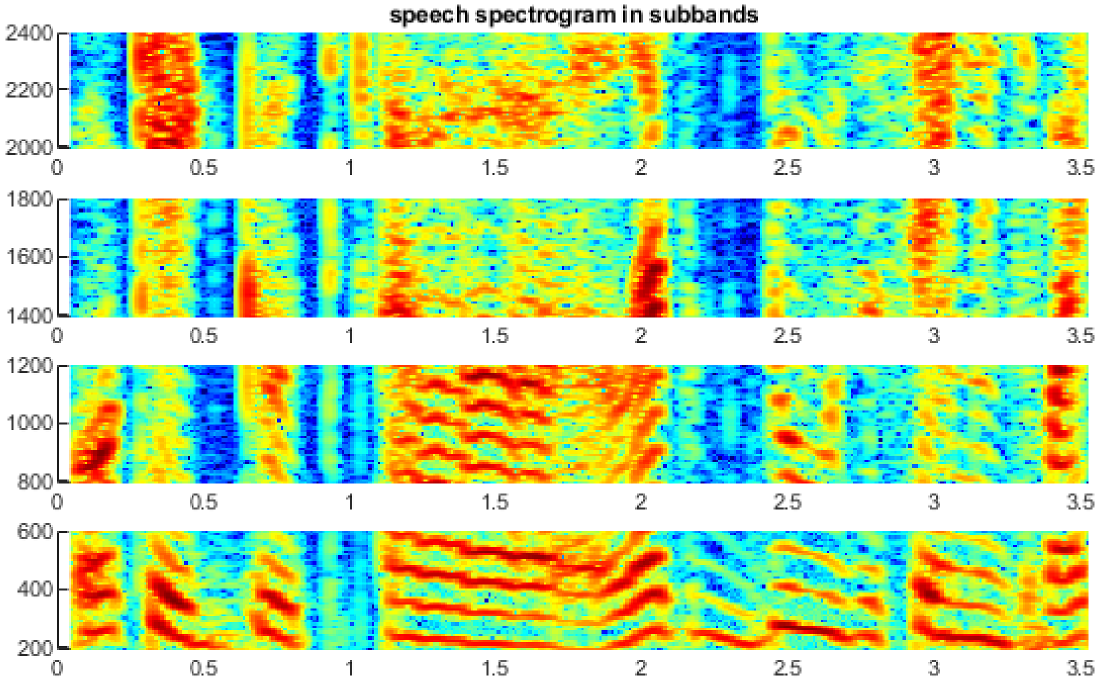
MIPS are dropping about o(1/M), and M is limited by the algorithmic latency rather than by optimal MIPS. The figure below shows basic MIPS for 5ms block size, which ends up (often) in 25ms algorithmic delay. As before, each LMS model would take 3x. For smaller blocks, you need to increase the MIPS inversely, for larger – decrease proportionally. Two model LMS-based AEC would take 10x, which is only 300MIPS, plus 30% control and FFT overhead – reasonable for SIMD devices nowadays.



The eigenvalue spectrum spread, leading to LMS convergence only on the most powerful harmonic / formant is addressed in a ‘divide and conquer’ style.



First, let's look at a typical speech spectrogram. We see that the spectrum is dominated by the 500+-200Hz region, and there is not enough energy in other frequency bands to converge upon.



Then, let’s split the spectrogram into 200Hz (1 / 5ms) bands, resample them properly and normalize each independently. Here we make an assumption of an ideal brick wall filtering – for the sake of simplicity only. We see that energy-wise it becomes quite possible to converge in each sub-band, the narrower a sub-band is, the less effect has the singularity of the source signal. Moreover, the fluctuating pitch creates chirp-like signal in sub-sampled subbands, and we know that chirp-like signal is full-spectrum, which is a very good excitation for successful LMS convergence.

SAF also allowed application of more elaborate approaches to solve echo path non-stationarity and abrupt echo path change detection. The field started to bloom. In 1995, Dennis R. Morgan and James C. Thi introduced open-loop delayless SAF, etc.

On the negative side of SAF, the sensitivity to loudspeaker’s nonlinearity actually became worse - we’ll discuss it later.

AEC, ANC, AFC (for hearing aids) became a reality. However, the performance was mediocre, much worse than anticipated, and reasons unclear. Problems persisted unsolved and performance targets remained unachievable. Since the early 2000s, the number of publications has started to decline. Alas, the majority of field applications were of horribly inadequate quality, regardless of the marketing efforts. The normal reaction to responding to a call using a speakerphone remained the same “Switch to the handset, please”.

6.2 Deficiencies

The core problems of SAF are:

1. Lack of understanding that for correct representation in the digital domain, the system to be identified MUST be band-limited, same as the I/O signals.

   1.1. The signals’ limitation of band-limited-ness is described in practically all DSP textbooks. However, it’s not discussed there sufficiently, and that does not translate into the real understanding. This understanding is reserved mostly to the people designing ADC / DAC. Unfortunately, all of the DSP books I looked at miss the requirement for the system identification problem to be also properly band-limited. Even the unquestionable leading experts in the field, from Sweden's school of adaptive control, have made such mistakes.

   1.2. As a consequence, the same filterbank for IN and OUT is used. Solution: IN filter must be sufficiently wider than OUT. The details are not trivial and will be explained later.

   1.3. As a consequence, there is a lack of understanding that in-subband adaptive processing is principally different from the traditional. The parallel and consecutive architectures (TBD) are different in essence.

2. Lack of understanding of the latencies involved in the processing. Solution: understanding of DSF (Delta-function Spread Function) and the need to augment sub-band RIR representations with room to spread, for t<0 (non-casual part) and after the end, t > LADF.

3. Lack of understanding of the implications of filterbank on the robustness of adaptive filtering.

   3.1. The SAF troubles do not happen if and only if all of the RIR’s spikes fall exactly on the integer multiple of the sampling period (fractional delay is 0.0) because fractional delay sinc() function becomes sampled as a delta function. Then, any variations in RIR time shifts will lead to very inconsistent convergence.

3.2. RLS is not the remedy. It can greatly improve (thanks to the embedded Gramm - Schmidt orthogonalization) frequency-domain identification, very close to the band-edge 0/0 singularity, but RLS does it on account of approaching sinc() sidelobes in the time domain. Thus, RLS easily produces the weirdest artifacts you can imagine, with combinations of complex biases and ripples which can bewilder you for months.

3.3. While the general character of the artifacts is predictable and repetitive, their specific waveform, either in frequency or time domain, are not repeatable at all, and slight variations in excitation may affect the artifact's waveform drastically

3.4. Moreover, give RLS enough time, and it will (and shall) diverge. The step matrix (or Kalman gain matrix) is an inverted Fisher Information Matrix. If RLS is run long enough, it will become a spectral filter, inverse to the input's Power Spectral Density (PSD) which is the IN spectrum squared (^2). The band-edges become untouched while the band's centre becomes attenuated. The relative power in the aliasing artifacts is the worst at the band's edge. RLS is not aware of the aliasing, and it shall and will diverge, sooner or later, when the band-edge aliased signal goes up and above in-band signal's residual error. More on that in Part III.

3.5. Applying subband filtering prior to band-pass sub-sampling diminishes the "equalizing" effect of the excitation's spectrum. DFT, being an orthogonal transform, does not affect the eigenvalue distribution. If we have a relatively flat-spectrum per-sub-band input signal, then, after subband filter-bank processing, it could not have the eigenvalue distribution any better than the prototype filter's shape. That is also the essence of the "band-edge" effect.

3.6. The lack of per-band band-edge convergence may pass unnoticed in full-band for an indiscriminate eye because it will be somewhat attenuated inside synthesis filterbank

3.7. Solution: you absolutely need to understand very well how adaptive algorithms work, what to test for, when and how.

4. Lack of understanding how to convert sub-band RIR estimates into full-band RIR estimation with required precision.

4.1. The existing algorithms for open loop delayless subband adaptive processing are not to the point and unnecessary complicated.

4.2. Due to the very slow decay of sync() side-lobes, the search for a finite-window based inverse transform (from sub-band into full-band impulse response), existence of which was assumed in open-loop delay-less SAF variations, has been of highly debatable nature. I mean we could find it in-potentia / in-silico but could not converge to it in-vivo.

4.3. Solution: there is no difference whatsoever between designing a bi-orthogonal synthesis filter to a given analysis filter and designing a RIR-per-subband-into-fullband synthesis filter, bi-orthogonal to the above-mentioned DSF.

5. Lack of understanding of RIR variability's auto-correlation matrix and how it affects the convergence, as well as the lack of general understanding of what adaptive processing is and is not, and how it works.

5.1. LMS is based on the implicit assumption that the initial covariance matrix of estimation errors (and non-stationary perturbations) is ~ flat, uniform along the diagonal i.e., the echo tail variations have delay-invariant distribution. This is definitely not true for acoustics… and I suspect it is not true for any stable system: IR must be decaying, to *norm(IR(1:T))* < ∞ for ∀*T*.

5.2. The exponentially windowed versions of LMS have been proposed, but they are stationary and miss quite a few the important details of LMS (and RLS) convergence.

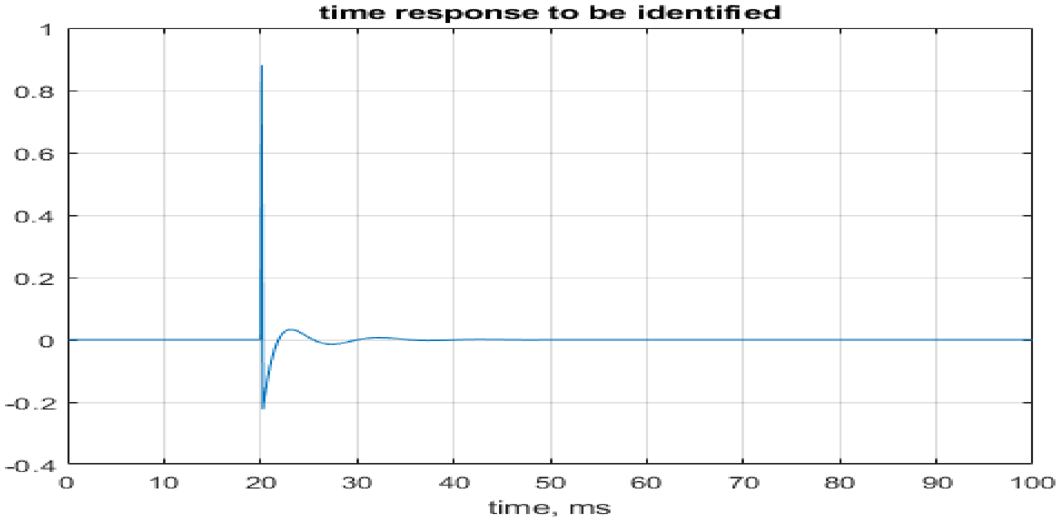
5.3. See Part II for the solution's directions and further details.

SAF-1998 was a first step in a right direction; a formulation of a challenge rather than its solution.

6.2.1 Non-band-limited-ness deficiency details [105]

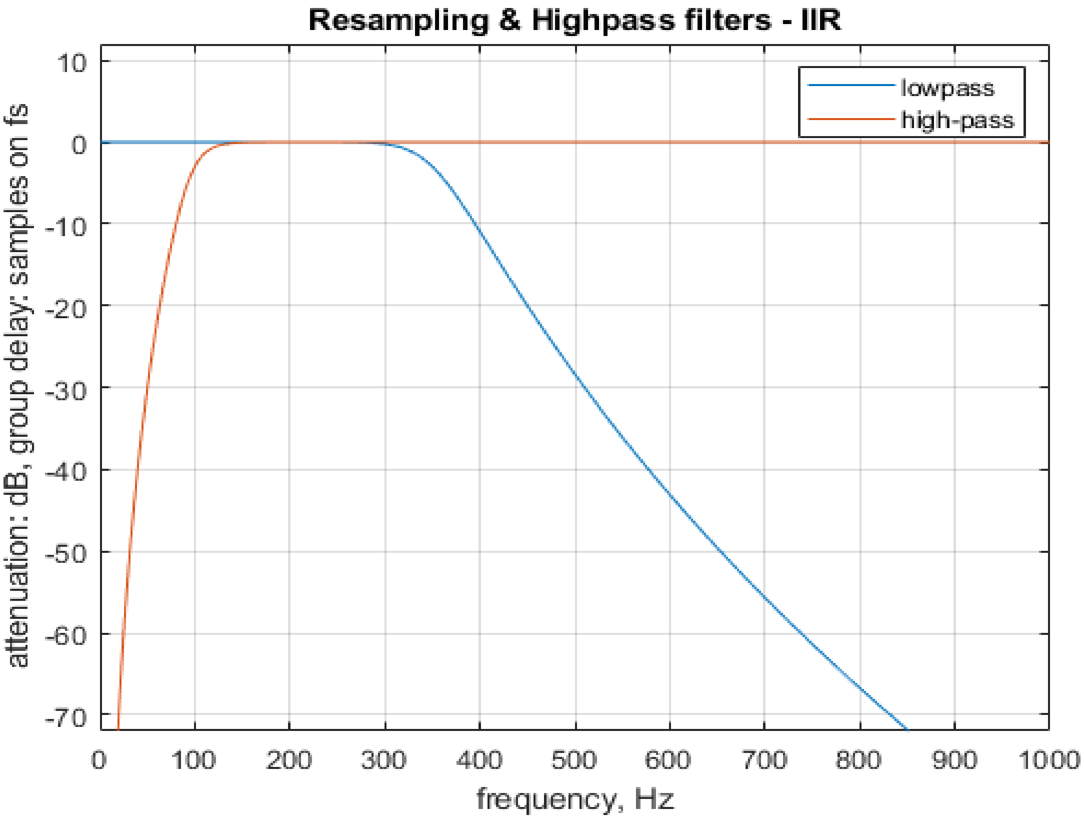
It may not be so easy to explain, so let's start from the basics.

- Let's suppose we want to implement an adaptive filter, with a sampling rate of fs=1kHz.
- Let's augment the adaptive filter with a few ms of "negative" time. We know that the real systems are casual, therefore this augmented part shall read 0s, ideally. In reality, it will have some "noise" which should be representative of Mean Square Error (MSE) of the entire adaptive filter[12].
- Let's simulate this approach on a simple system response, such as 100Hz High-pass filter, delayed by 20ms. [doc_p105.m]



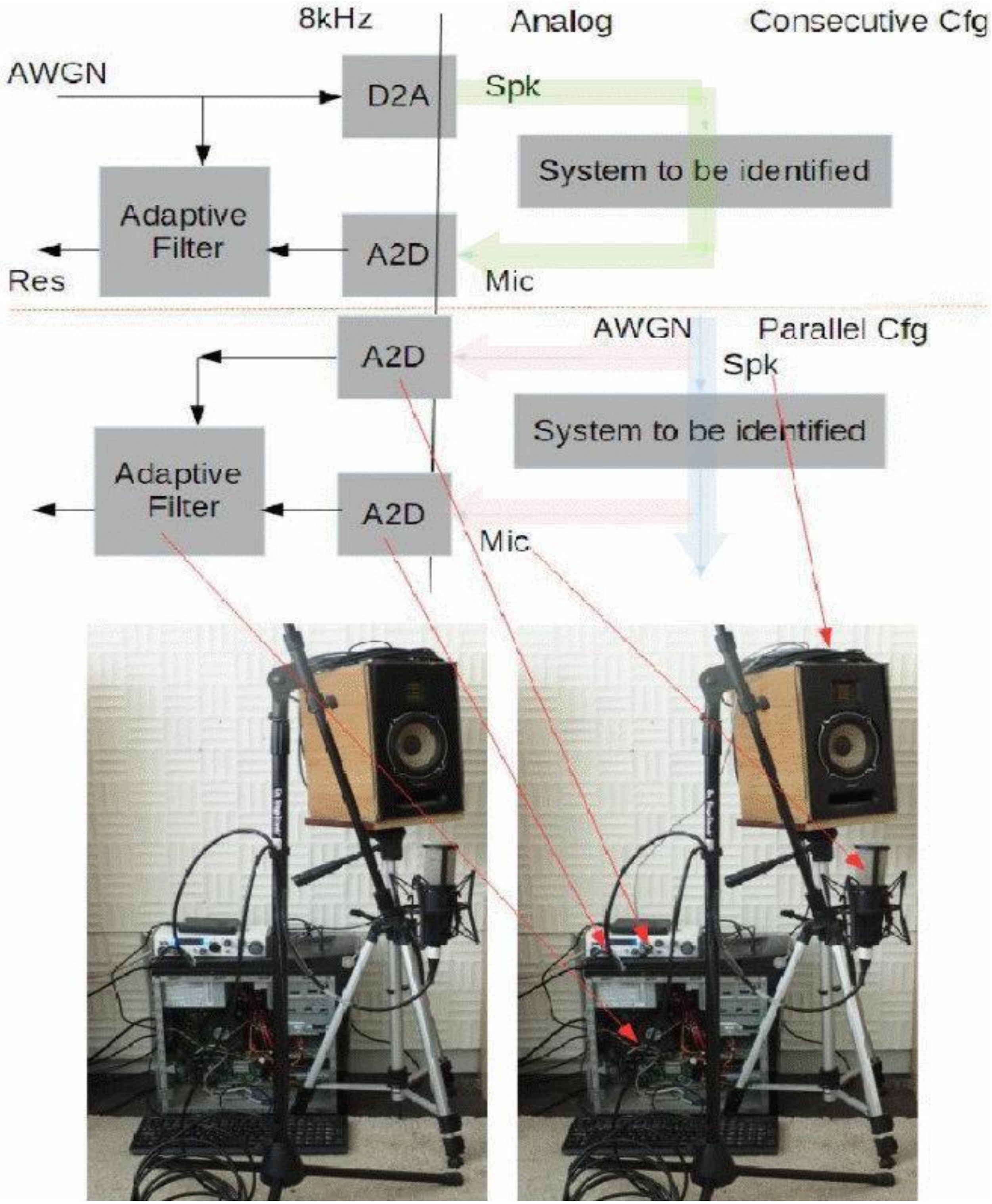
- Let's simulate analog by 8x higher frequency, FS=8*fs.
    - Let's simulate Digital to Analog converter (DAC) by upsampling to high frequency and a low-pass filter to remove aliases.
    - Let's simulate Analog to Digital converter (ADC) by a low-pass resampling filter followed by decimation to fs=1kHz.
    - Let's use 9th order Butterworth IIR as a low-pass resampling filter.

[12] This idea was floating around for quite some time (at least early 80s), and many researched independently re-invented it.



6.2.2 Consecutive vs Parallel ADF Configurations

Two configurations are considered (see the picture below):

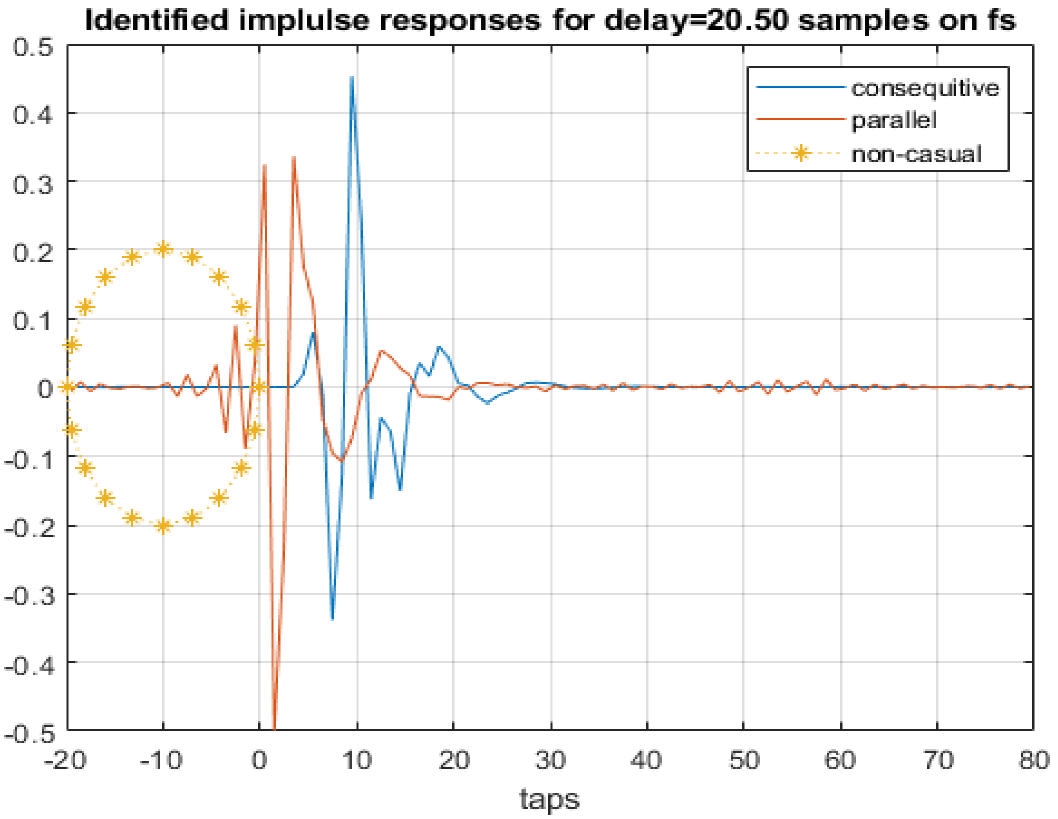


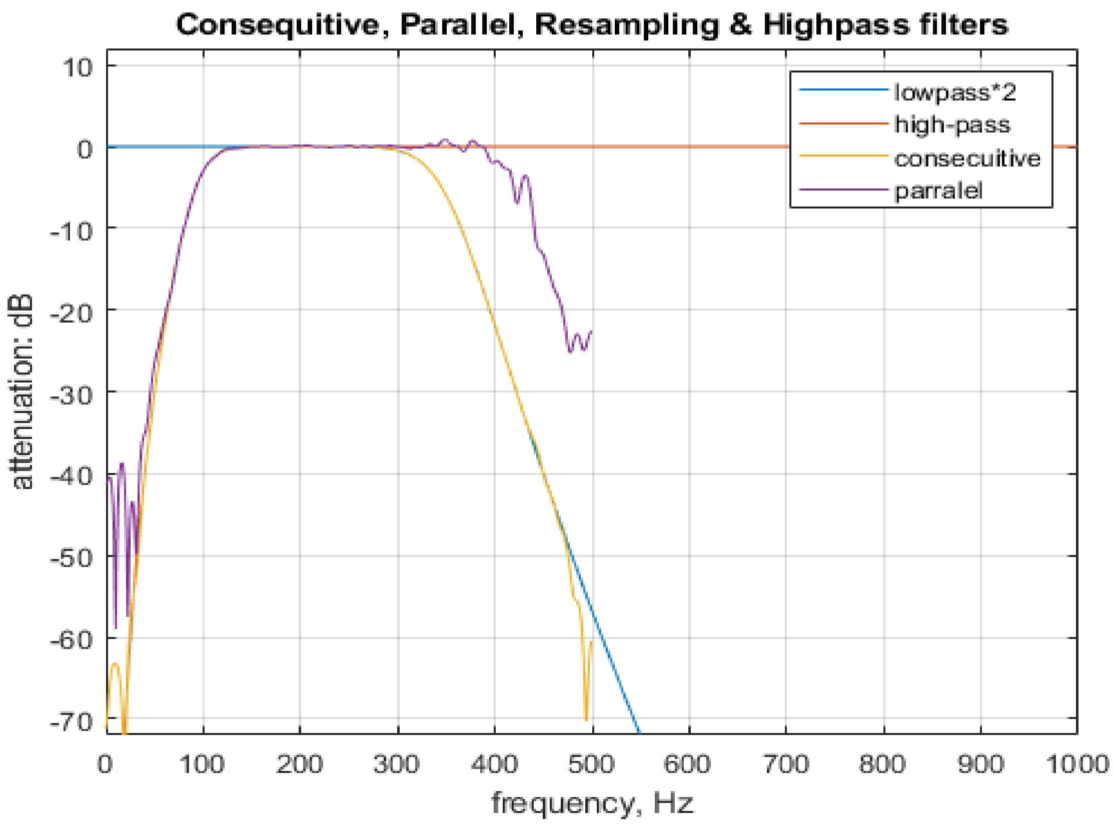
1. Consecutive ("classical") configuration, when digital, 1kHz, “left” side creates excitation, which is then converted to analog, goes through the system to be identified, the output is digitized and adaptive filter is run.

2. Parallel configuration, when both the input and output of the system to be identified are sampled on fs=1kHz, and digitized, and the same adaptive filter is run.

The main "physical" difference is the addition of gray RCA -> XLR cable which is sensing the input to the audio monitor using an "Y" splitter (not shown) and sending it to the second input of audio interface (the used loudspeaker, a modified F5, has HPF of about 80Hz).

- The excitation is 1 second of white Gaussian noise in both cases, which should be plenty if we want to identify only 30...40ms of the simulated system response.
- The adaptive filter is Kaczmarz's (which is often called [N]LMS) with step size set to 1.0, as for noiseless [mic] case.

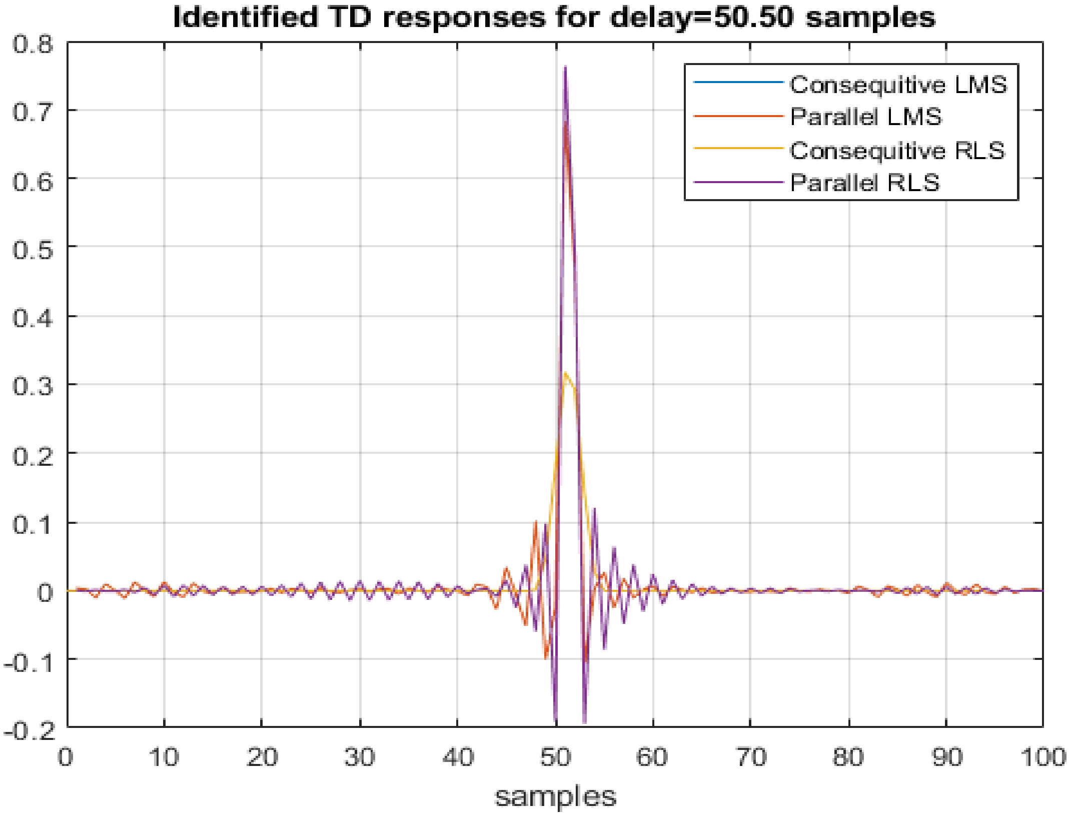




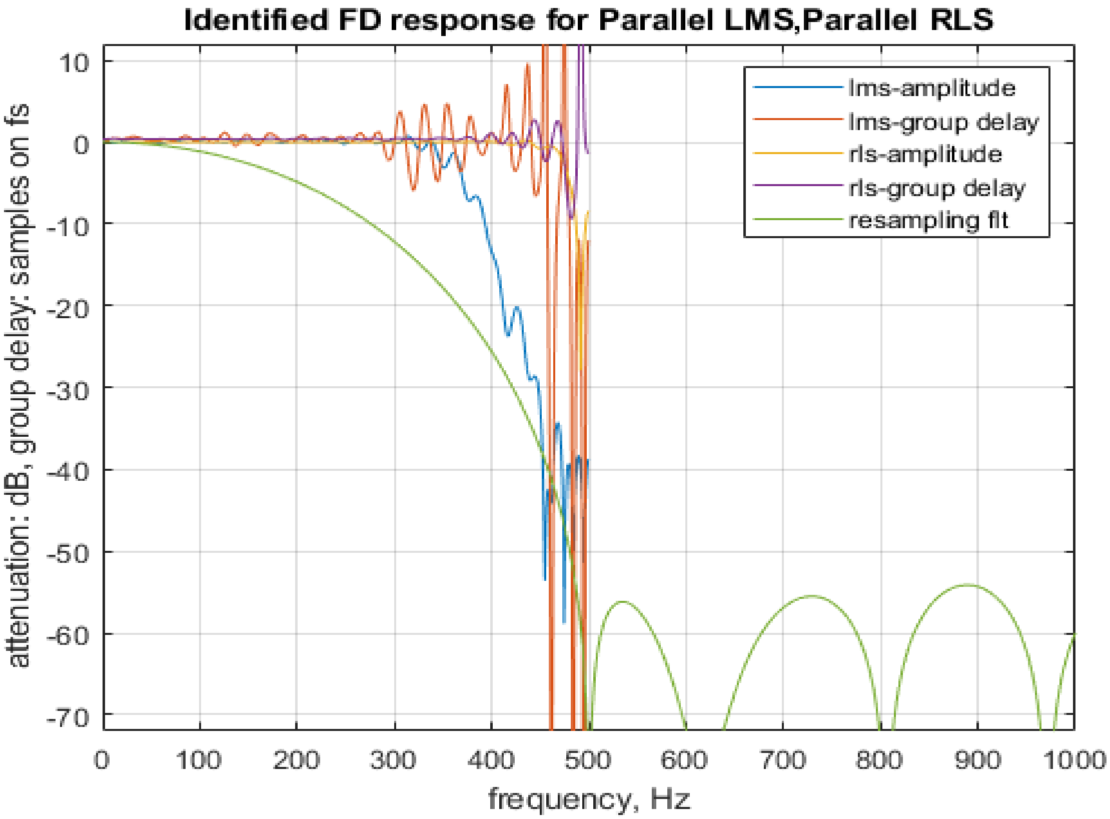
- The impulse response identified by the consecutive configuration is meaningful but by the parallel configuration is confusing... so much that you may even start looking for a bug:-)
- For consecutive configuration, you can see that LPF IIRs inserted some group delay, and the tail is a convolution of HPF and in/out anti-aliasing LPF IIRs (which decays much faster, of course).
- For parallel configuration, you can see the same HPF tail, but fs/2 ripples are meaningless, especially their extension into non-causal time [-20... 0].
- But this is not a software bug. This is “the” problem in the SAF-1988 architecture.

Let’s simplify the system to be identified even more.

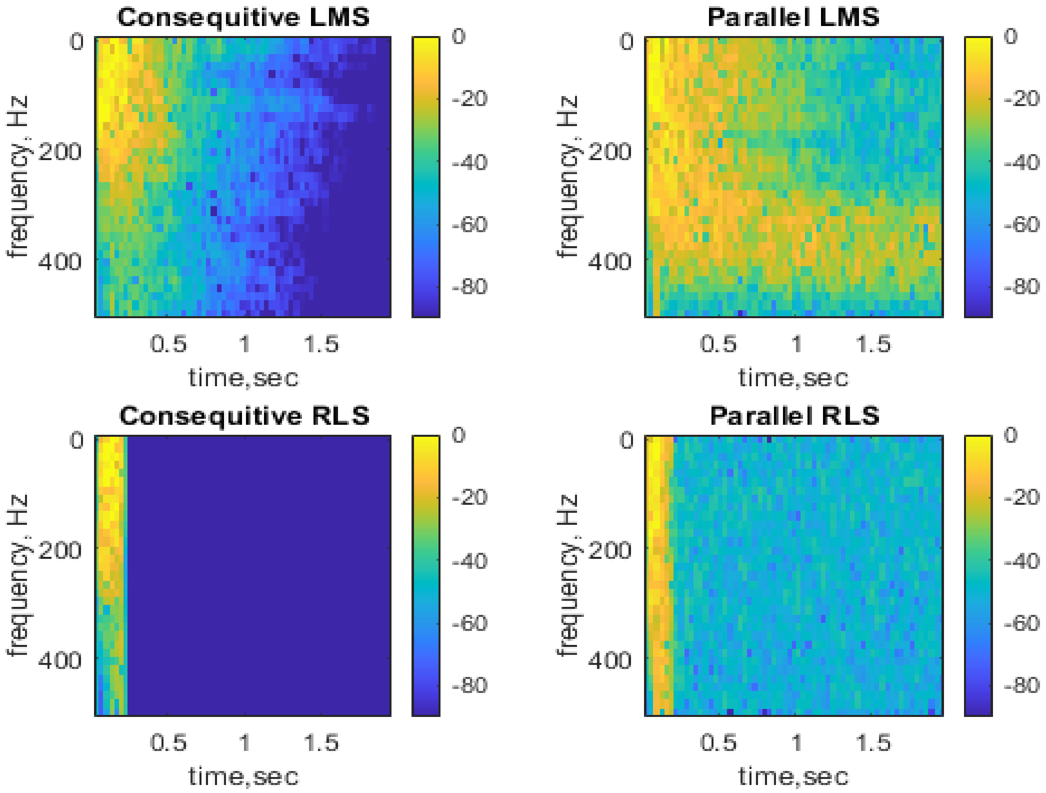
- No HPF,
- pure delay only, for about 50 samples (right in the middle of impulse response).
- Instead of IIR LPF anti-aliasing filters, let's use a short FIR, 6 samples on fs - so that we know for sure their effects may not stretch all over the identified impulse response.
- Both LMS and RLS are used [doc_p106.m].



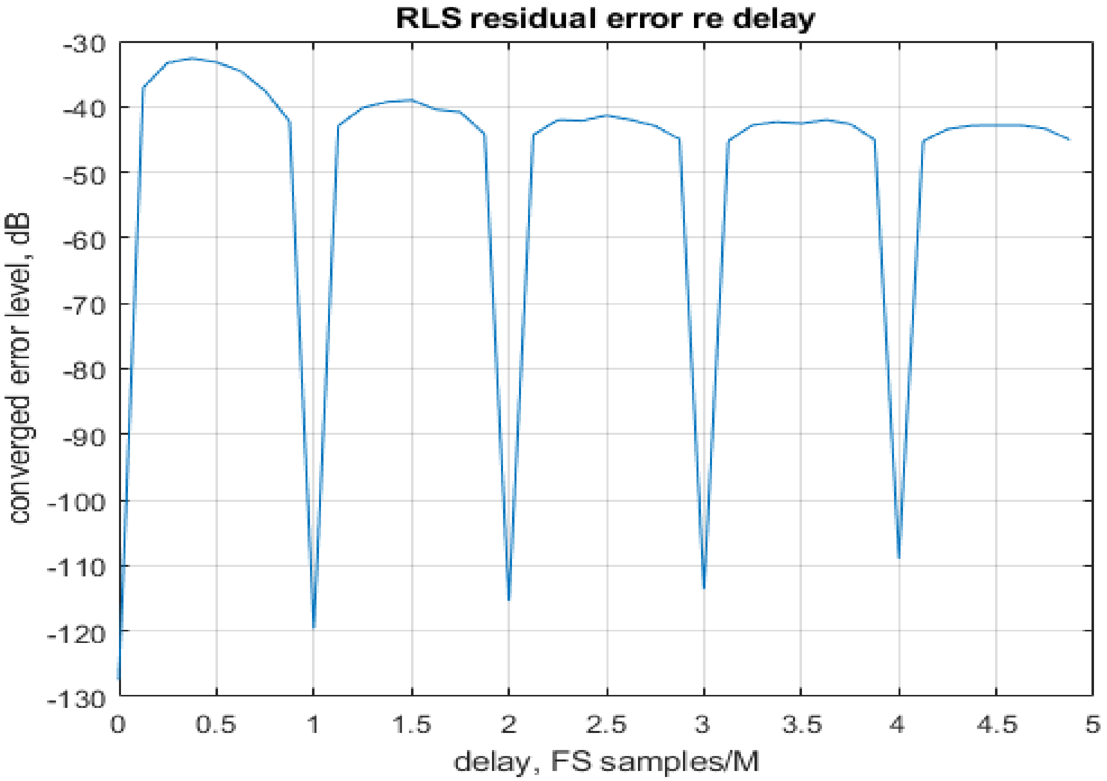
- So, we can easily see that parallel configuration is principally different from classical. It is not "forget all you learned before" but ... close.
- ~fs/2 ripples still stretch all over the impulse response, both in casual and non-causal direction, both for LMS and RLS.



- The high-frequency part of the system's frequency response is not really identifiable, for both LMS and RLS.



- The long-term low-frequency convergence of LMS adaptive filters is constrained by “band-edge effect” a.k.a. (lack of) high-frequency convergence, due to the LMS Spectral Defect, discussed above.
- We also can see that using LMS is inherently a very poor idea for parallel configurations. LMS is well suited to converging on well- defined flat-spectrum signals, same as the fastest gradient is well suited to minimization of well-defined quadratic forms. Give it a more difficult task, and it would stall indefinitely. But a parallel configuration is inherently a not-well-defined problem because of the low-pass re-sampling filter.



- Moreover, if we vary delay from 0 to, say, 5 down-sampling ratios, we see that the performance of parallel RLS (and LMS too) is highly inconsistent. On delays = M*k, it “works” but on others – no, and the closer to 0, the worse.

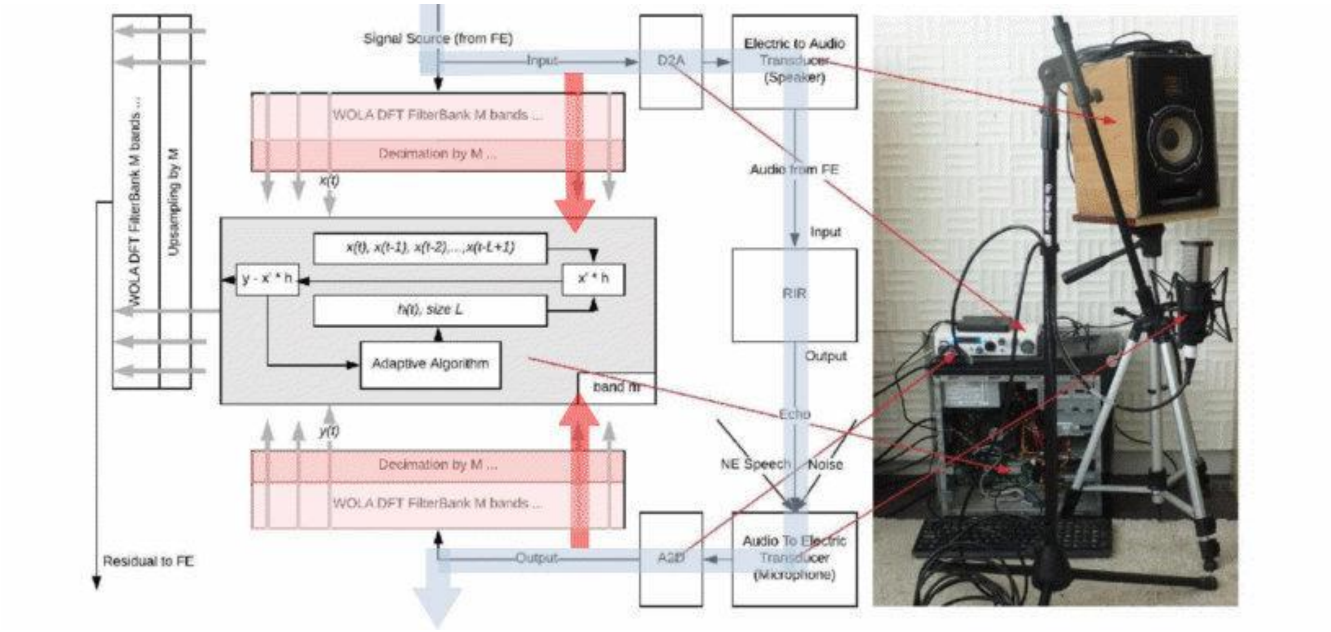
Now we understand that all those troubles are real and should be expected because we attempted to identify an ideal fractional response filter (input filter / output filter == 1 for all frequencies <= 1kHz), which is sinc() in the “analog” time domain.

The system to be identified should be represent-able in the digital domain, i.e. at fs/2 the system frequency response (all all it's derivative) shall be as close to 0 as possible, so that the sidelobes' (time-wise) amplitude is low and decay is fast, because any discontinuity in f() or its derivatives leads to proportionate sidelobes with 1/f^n decay where n is the derivative order.

The real systems rarely if ever possess such quality, so they must be conditioned appropriately - pretty much the same as the input analog signals to DSP are conditioned in A2Ds. In the consecutive case we simply get lucky because the D2A and A2D not only condition the signals but the system identified as well. In the parallel configuration, nobody is doing such conditioning for us.

6.2.3 Relation between parallel configuration and SAF

We could have dismissed all these nuances and ignored them, if the parallel configuration were not the exact configuration for each per-subband adaptive filter, operating in each sub-band independently.



Here, the full-band frequency serves as "pseudo-analog" quasi-sampling frequency, and the subband sampling frequency serves as fs=1kHz.

7 Overcoming deficiencies of SAF-1988 by FSAF

This set of techniques addressing these problems is called Fast Subband Adaptive Filtering (FSAF). They are the subject of this discussion, of this MATLAB package, and of this book describing it.

7.1 FSAF Operating Principles

The first technique is based on noticing that there is no need whatsoever to make the IN filterbank identical to the OUT filterbank. The prototype for Input filterbank does not have to be the same, nor the same length, nor symmetric. It's required that IN is "fuller" than OUT so that the sub-band system response is sufficiently band-limited and that IN is not ruining eigen-spectrum as badly as a usual OUT filterbank.

Other techniques are related to the internals of adaptive algorithms, and they must be used together.

7.2 Deficiencies

TBD by you, as well as how to overcome them :-)

8 Book Overview

Part I - Introduction

This part of the book is an introduction, a short historical review and the problem statement.

Part II - Adaptation

Discussions in this part are centered on delta-function RIR and white noise excitation

- The system to be identified must be band-limited

- Consecutive is not the same as Parallel, in depth
- The properties of RLS and LMS which are important for audio applications of subband adaptive filtering
- Modifications to account for under-modelling
- Introduces the Diagonal Least Square (DLS) algorithm and its variants (shelf-DLS and block-DLS).
- Robustness of adaptive algorithms
- Introduces the meta-adaptive approach to improve robustness, which also allows a significant reduction of adaptive models without compromising the coverage of possible use cases
- Discussion of meta-RLS and meta-DLS algorithms.
- APA as dynamic spectrum equalization and where/when it is useful.

PART III - FILTERBANK

- FSAF filterbanks (for IN, OUT and RES) must satisfy several conditions which must be explicitly checked during design:
    - IN(f) filter is sufficiently wider than OUT(f), to suppress any and all deficiencies and byproducts of non-ideal band-limitedness.
    - OUT(f)./IN(f) Delta-function Spread Function (DSF) sidelobes are small and decay fast
    - Timing invariance: RLS / LMS convergence shall not be sensitive to delta-function offset
    - Additional delay due to DSF non-casual sidelobes.
    - RLS / LMS convergence speed and residual error spectrum are reasonable
    - DSF shall be robust to variations in method, under-modeling, noise, and other factors.
    - A reasonable synthesis filter shall exist for the chosen OUT(t) filter, and OUT -> RES reconstruction error shall be sufficiently low.
    - For a converged FSAF, the full-band noise spectrum is flat-ish.
- Open-Loop Delayless (OLD) FSAF, with RIR (nearly) Perfect Reconstruction (as per requested precision)
- Nested / Composite FSAF using Open-Loop Delayless FSAF
- Large values for subsampling ratio M are treated in a separate chapter.
- There is no “ideal” filter, optimal for all FSAF applications
- The regularizations of RLS class adaptive algorithms to prevent divergence due to aliasing

PART IV - APPLICATIONS, ISSUES AND NUANCES

This part discusses the real-life applications with real-life RIRs and real-life excitations:

- Framework

- AEC: echo canceller
- ARC: modification of AEC to control reverberation. Looking for zero-delay RIR inversions is not an undebatable approach.
- AFC: modification of ARC to control feedback
- ANC is not discussed in appropriate details. ANC is seriously limited by the acoustics, also ANC belongs to the domain of adaptive control, not of adaptive filtering, per se.
- System identification on the example of loudspeaker-room impulse response.

Numerous implementation nuances, typical mistakes, pitfalls, known deficiencies and still open problems that need to be researched, etc are also discussed here.

Part V - Manual

This part is the manual to the classes and functions.