6

Acoustic Interfaces Exploiting Sparsity Constraints: An Experimental Study

Contents

6.1 AEC experimental conditions **80**

6.2 Performance advantages of proportionate filters **82**

6.2.1 Simplest scenario: exact path modelling in absence of near-end speech 82

6.2.2 Exact modelling scenario in adverse environment . 86

6.3 Performance analysis of VSS proportionate filters **88**

6.3.1 Under-modelling the acoustic impulse response . . 89

6.3.2 Robustness against double talk 91

THE objective of this chapter is to present by means of simulations the most important features of the proportionate adaptive algorithms described in the previous chapter. In particular, we analyzed the behaviour of those algorithms in several conditions and we investigate the performance of the proposed variations in order to give an overall description

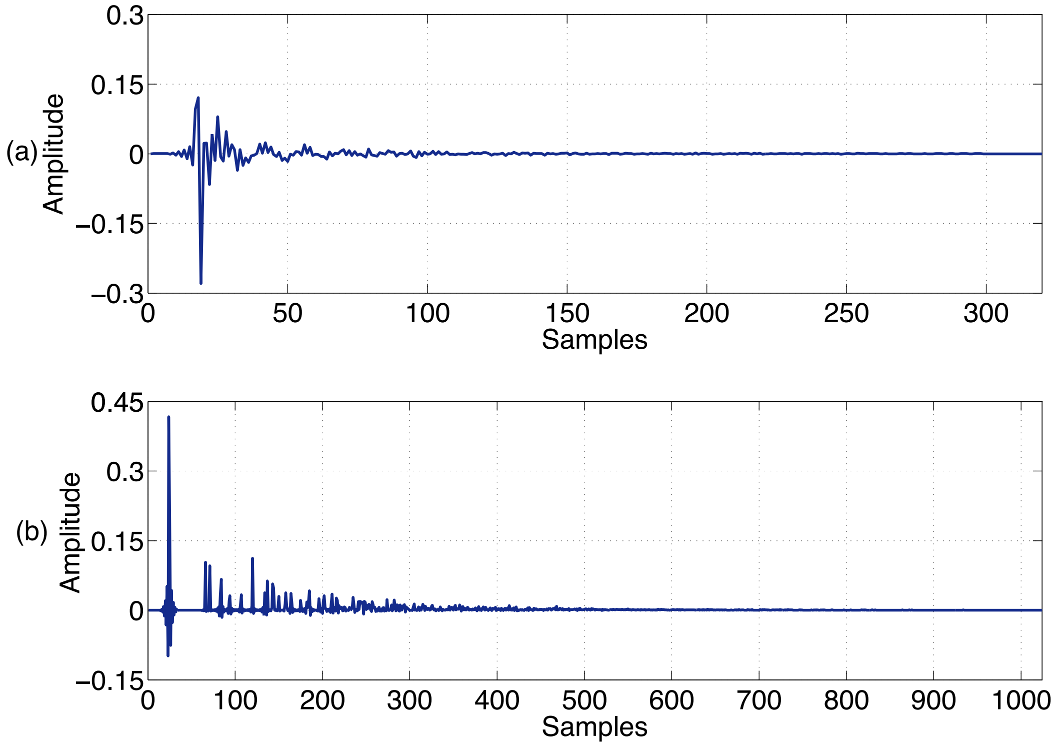

Fig. 6.1: *Acoustic impulse responses used in simulations. (a) Real AIR measured in a low reverberant room. (b) Simulated AIR with a reverberation time of 130 ms.*

of the effectiveness of proportionate algorithms. Most of the experiments are conducted in acoustic echo cancellation scenarios, which allows to better comprehend the capabilities of the algorithms.

6.1 AEC EXPERIMENTAL CONDITIONS

In this first part of the chapter we show experiments conducted in the context of echo cancellation since it is the best acoustic application to evaluate the effectiveness of an adaptive algorithm.

Experimental simulations in an exact modelling case were performed using a real echo path measured using a low-cost loudspeaker inside a room with short reverberation time. This AIR is composed of 320 coefficients and it is depicted in Fig. 6.1 (a). When we have considered an under-modelling scenario, experiments have been conducted using a different AIR, simulated by means of a Matlab tool, *Roomsim* [24], and is measured by using an 8 kHz sampling rate. This simulated AIR has been achieved considering a $(10 \times 6, 6 \times 3)$ m room with a reverberation time of $T_{60} \approx 130$ ms. It consists of 1024 coefficients; however when we consider an under-modelling filter we truncate it after the first 512 coefficients. The simulated AIR is depicted in its total length in Fig. 6.1 (b).

The far-end signal, i.e. the input signal, is either a white Gaussian noise signal or a female speech signal. The output of the echo path is corrupted by an independent white Gaussian noise (which simulates the near-end background noise) providing a *signal-to-noise ratio* (SNR) of 20 dB. All the signals are evaluated over a length of 10 seconds. Most of the simulations are conducted in a single-talk case, i.e. in absence of near-end speech input; however, we also use a double-talk scenario to evaluate VSS-based algorithms.

In addition, we want to prove the effectiveness of the algorithms even in adverse environment conditions, in which the acoustic environment changes due to a nonstationary source or to an alteration in the environmental conditions. In order to introduce an abrupt change in the acoustic environment we shift the AIR circularly to the right by 20 samples, 5 seconds after the start of the adaptive process.

In order to have a fair comparison we use, where possible, the same parameter setting for all the algorithms. Performance are evaluated in terms of *normalized misalignment* and in many cases also in terms of *ERLE* (see Section 3.4).

6.2 PERFORMANCE ADVANTAGES OF PROPORTIONATE FILTERS

6.2.1 Simplest scenario: exact path modelling in absence of near-end speech

In the first set of experiments, we evaluate the performance of proportionate algorithms with respect to the correspondent classic ones. We start our analysis taking into account the simplest algorithms (having unitary projection order) introduced in Chapter 5, i.e. the *normalized least mean squares* (NLMS) (5.16) and its proportionate version that we denote as IPNLMS, as its original indication [13]. We consider an exact modelling scenario in absence of near-end speech; the AIR used for these simulations is the one represented in Fig. 6.1 (a). We use the same parameter setting for both the algorithms: a step size

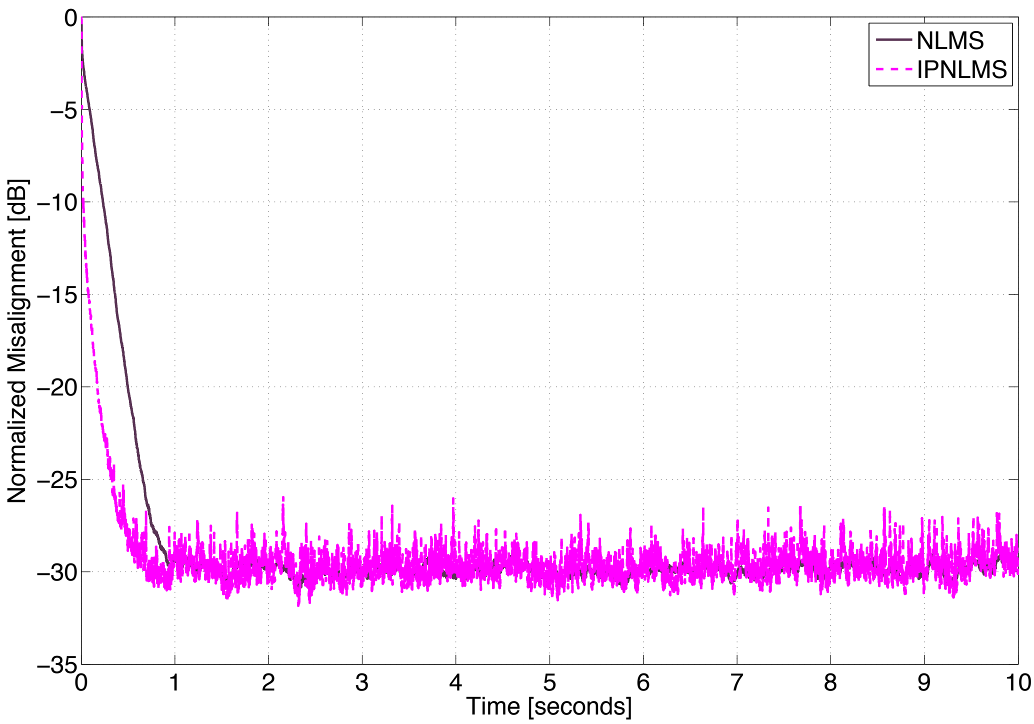

Fig. 6.2: *Misalignment of NLMS and IPNLMS algorithms with a white Gaussian noise input.*

value $\mu = 0.2$ and a proportionality factor of $\alpha = 0$; in addition, we choose a regularization parameter of $\delta_{\text{NLMS}} = 30\sigma_x^2$ for the NLMS, where σ_x^2 is the input signal variance, and a regularization parameter for IPNLMS δ_{p} according to (5.20). When the far-end signal is white Gaussian noise it is simple to certify a performance improvement of IPNLMS compared to NLMS in terms of convergence rate, as it is possible to see from the behaviour of the normalized misalignment in Fig. 6.2. The difference between NLMS and IPNLMS is more evident when the far-end signal is a speech input. Performance of IPNLMS are clearly improved in terms of filter misalignment, depicted in Fig. 6.3; moreover, an evident advantage results in the quantity of cancelled echo, i.e. in terms of ERLE, as it is possible to see in Fig. 6.4.

In Fig. 6.5 we evaluate the misalignment performance of a selection of PAPA algorithms with different projection order in case of speech input.

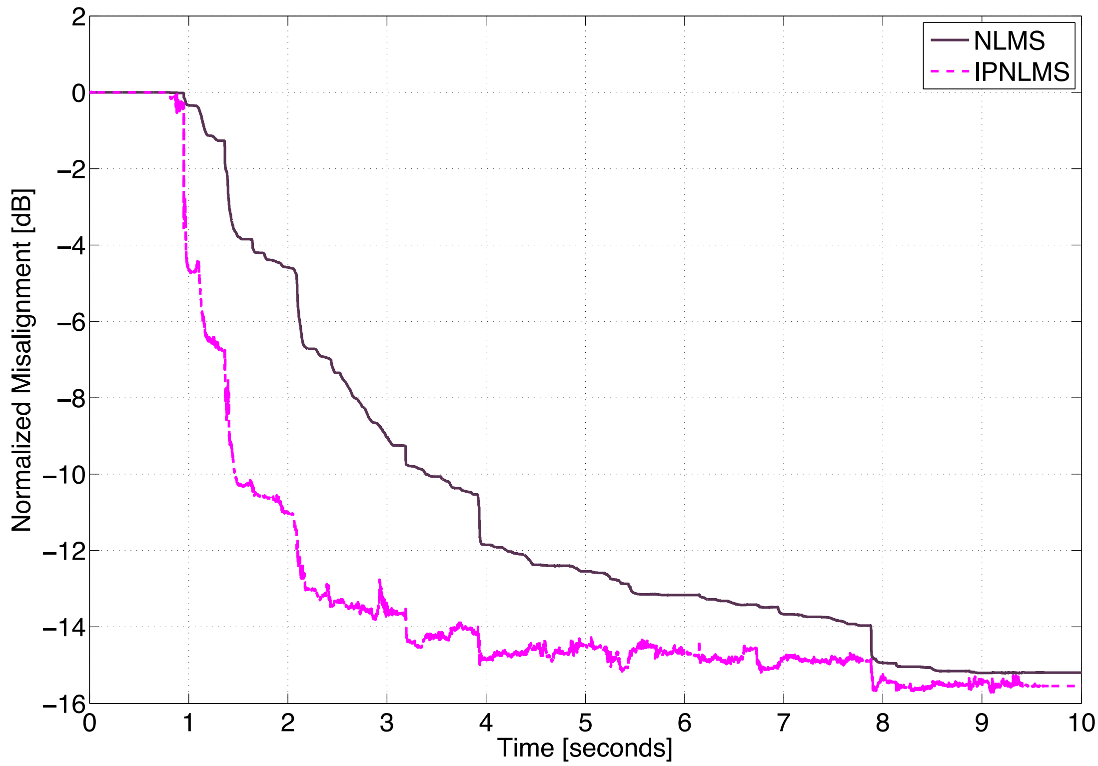

Fig. 6.3: *Misalignment of NLMS and IPNLMS algorithms with a female speech input.*

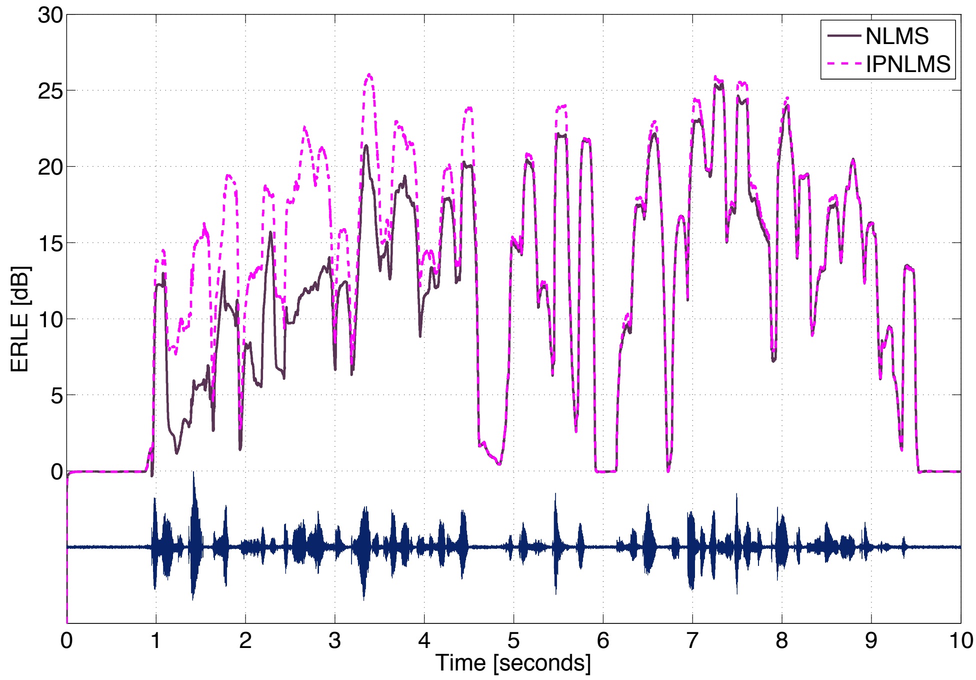

Fig. 6.4: *ERLE of NLMS and IPNLMS algorithms with a female speech input. The speech signal is reported for clearness.*

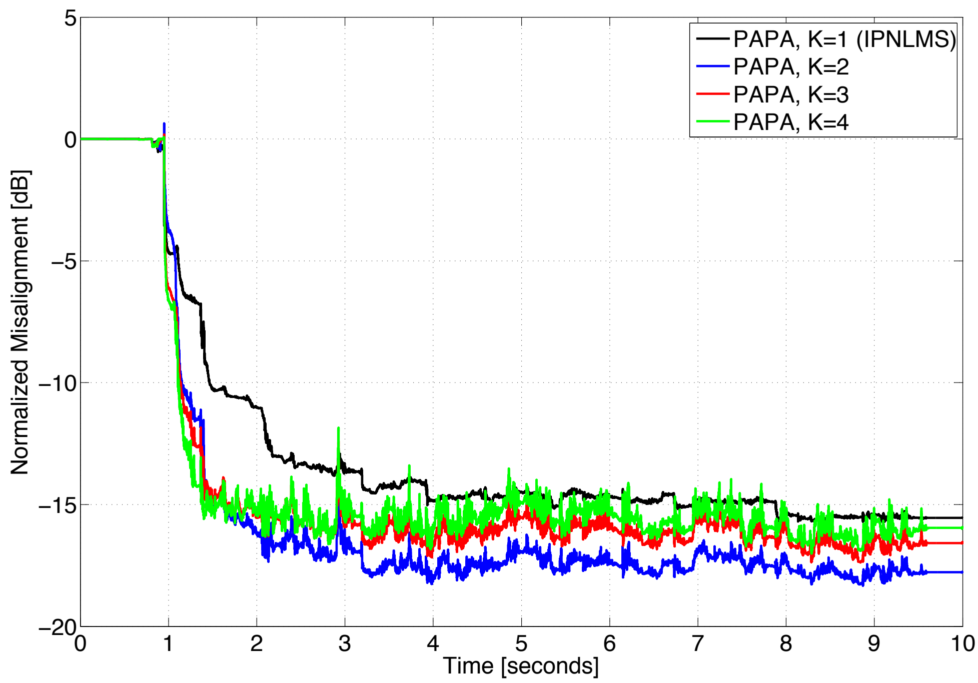

Fig. 6.5: *Misalignment comparison of PAPA algorithms with different projection order in case of female speech input.*

Let us note that in this case we evaluate only the speech input since the whitening capabilities of APA algorithms are obviously not evident when the input signal is already a white signal. From Fig. 6.5 we gather that satisfactory results can be obtained with a projection order equal to $K = 2$, or $K = 3$ at most.

We have also investigated the behaviour of the PBAPA (see Section 5.4). We report the comparison between PAPA and PBAPA in Fig. 6.6 in terms of filter misalignment when the input signal is speech. For both the algorithms we use a projection order of $K = 2$.

The behaviour of PBAPA misalignment confirms as said in Section 5.4: due to its structure the PBAPA overcomes PAPA misalignment at steady-state while showing poorer convergence performance. Due to this result we can say that PBAPA could be suited for applications with quite stationary conditions; however, if we consider AEC scenarios with adverse environment conditions

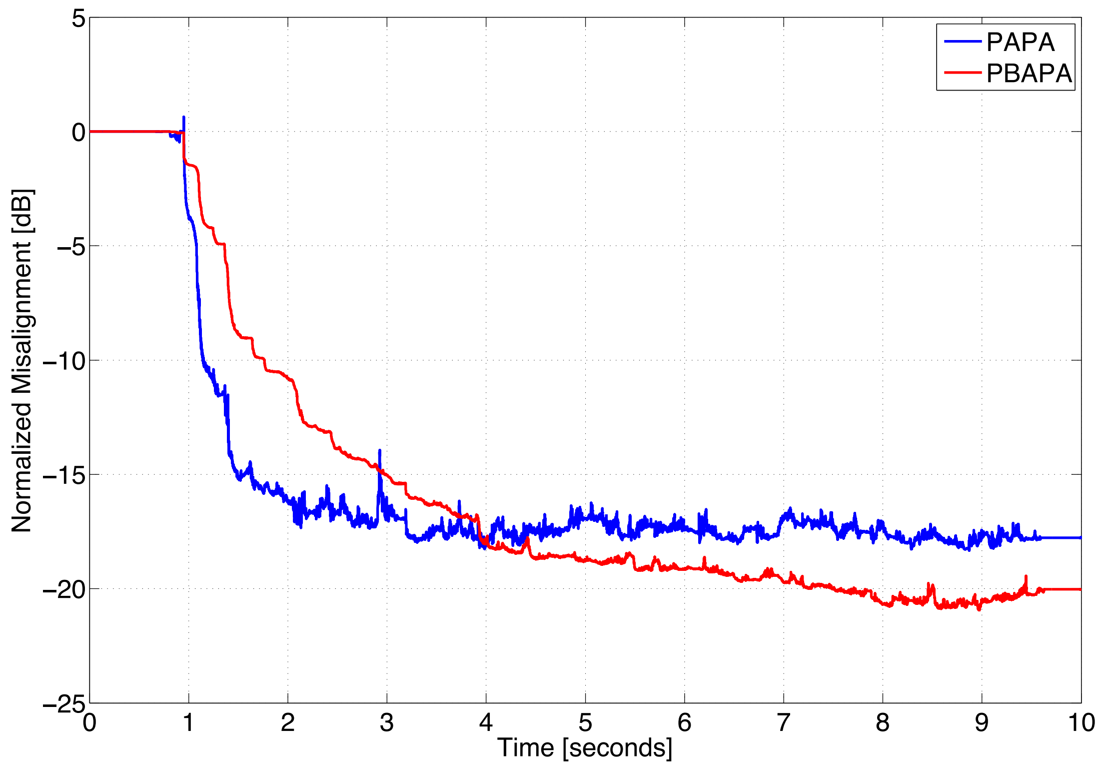

Fig. 6.6: *Misalignment comparison between PAPA and PBAPA algorithms.*

we still prefer the PAPA.

6.2.2 Exact modelling scenario in adverse environment

In this set of experiments we consider worse environment conditions respect to experiments conducted in the previous section. In a real AEC scenario several factors can be involved, thus altering the environment conditions, a source position change rather than an alteration of the environment temperature or the sudden presence of a new interfering source. When such an alteration occurs, the filter has to be readapted, so in order to achieve performance improvements an adaptive algorithm must have good tracking capabilities, i.e. a faster convergence rate in readapting.

We repeat some of the previous most representative experiments only

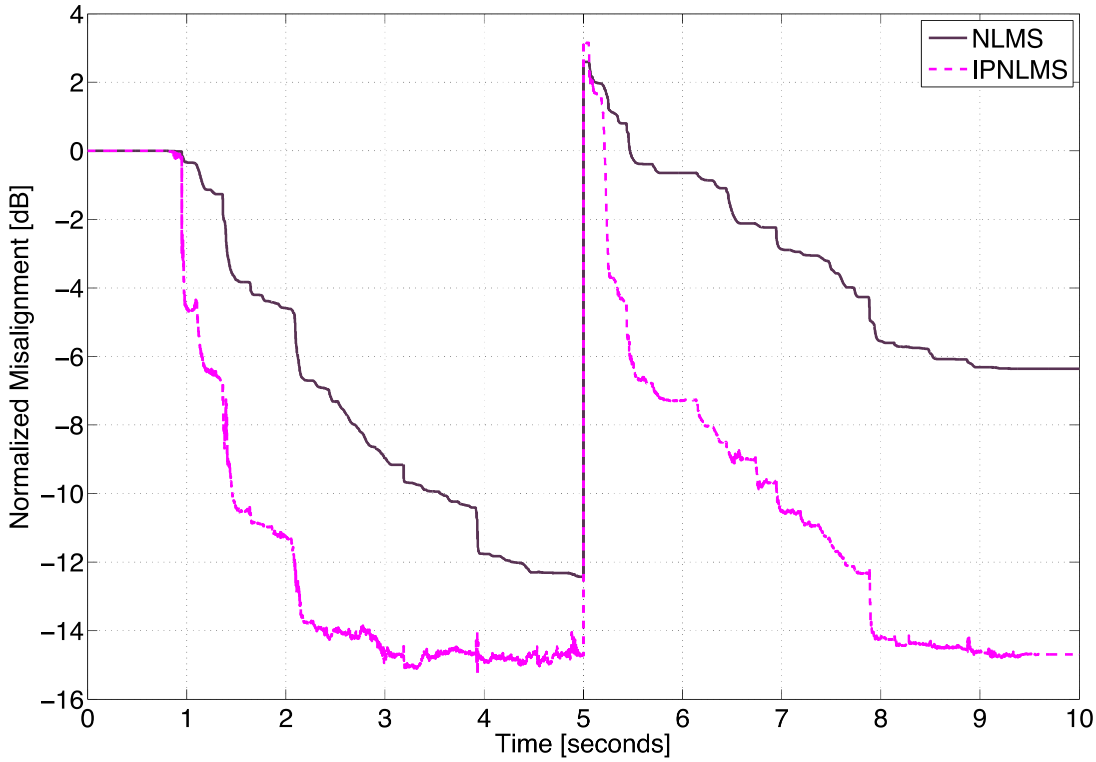

Fig. 6.7: *Misalignment comparison between NLMS and IPNLMS algorithms when a path change occurs. The far-end input is a female speech signal.*

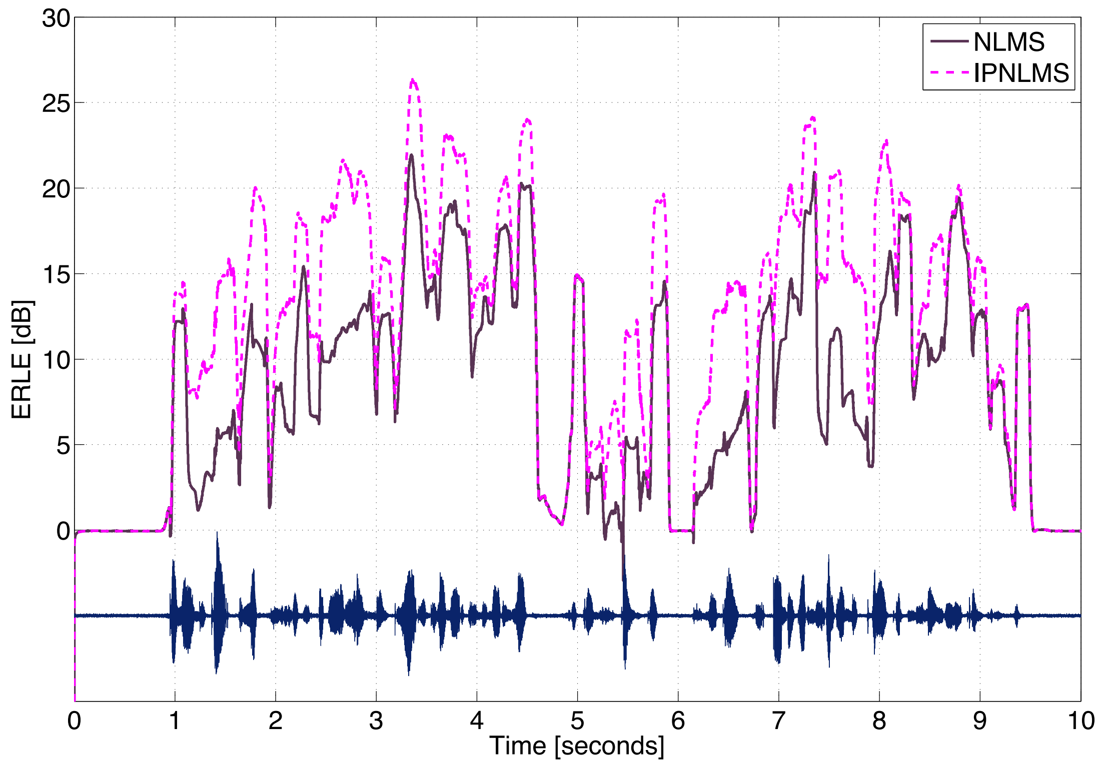

Fig. 6.8: *ERLE of NLMS and IPNLMS algorithms with a female speech input. The AIR changes at fifth second.*

changing the environment conditions, and in particular introducing a path change, due to an alteration in the environment, which occurs 5 seconds after the start of the adaptive process. In case of speech input it is possible to see in Fig. 6.7 that misalignment performance improvement of IPNLMS results more evident in adverse environment conditions compared to the simpler scenario in Fig 6.3. Comparing Fig. 6.4 and Fig. 6.8, when a path change occurs improvements even increase in terms of ERLE, since the behaviour of IPNLMS always keeps an advantage margin with respect to NLMS. It can be notice in Fig. 6.8 that the ERLE improvement (in dB) is directly proportional to the convergence rate; in fact, just after seconds 0 and 5, i.e. in transient state, the ERLE improvement is small due to a filter adaptation, while a larger improvement is achieved in steady-state, i.e. in time intervals $1.5 - 5$ and $6.5 - 10$ seconds.

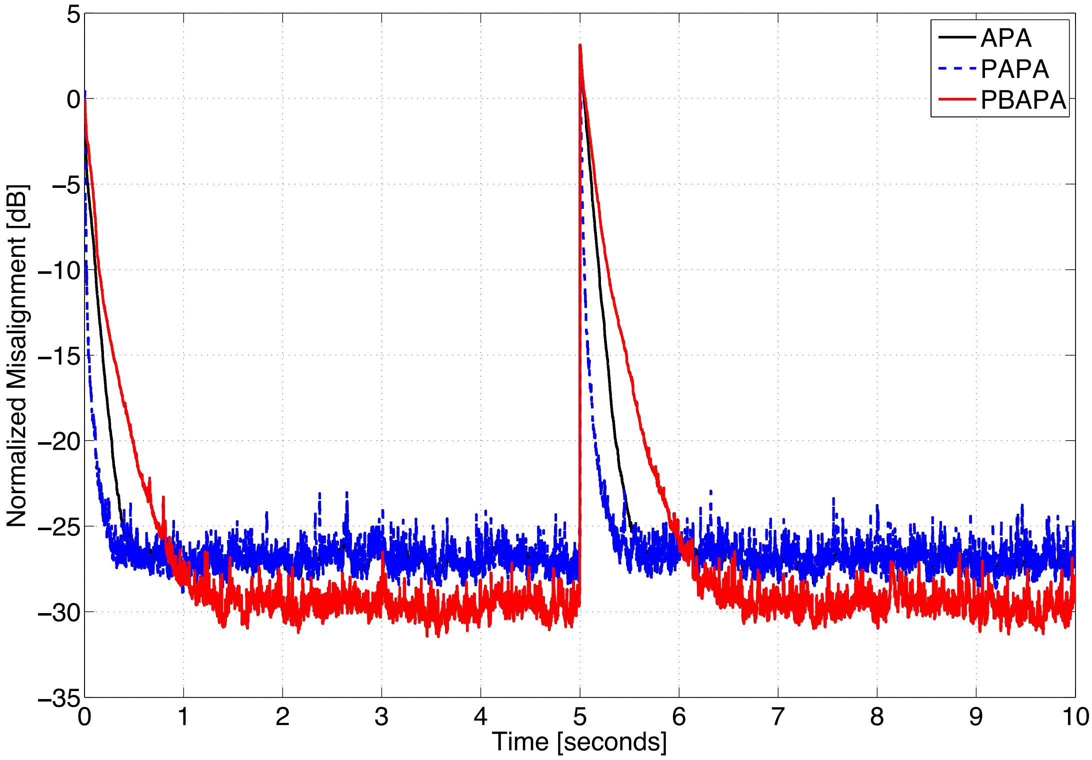

Fig. 6.9: *Misalignment comparison between PAPA and PBAPA algorithms with a white Gaussian noise input when the echo path changes. PBAPA shows better performance in steady-state; however, its tracking performance is poorer compared to IPAPA.*

We also investigates the behaviour of PAPA algorithms, including the PBAPA, when the echo path changes. Misalignment performance, depicted in Fig. 6.9, confirms the analysis done in the previous subsection, i.e. the PBAPBA provides the best steady-state behaviour while the PAPA shows the best tracking performance.

6.3 PERFORMANCE ANALYSIS OF VSS PROPORTIONATE FILTERS

Variable step size algorithms can bring significant improvement according to the environment conditions. In fact, due to their nature, VSS algorithms provides tracking performance improvements [124, 99] and this is the reason why VSS algorithms are well suited for AEC scenarios with adverse environment conditions and in presence of double talk. Moreover, VSS algorithms do not suffer from any under-modelling noise (see Section 5.5) and this allows to estimate the AIR with shorter length than the exact AIR length.

The variable step size approach introduced in Section 5.5 can be applied to any proportionate algorithm; however, for a performance analysis purpose we evaluate the behaviour of VSS-PAPA with a projection order of $K = 2$. For the set of experiments conducted in this section we use the AIR simulated in typical office room and depicted in Fig. 6.1 (b), whose length is $M_{\mathrm{A}} = 1024$.

6.3.1 Under-modelling the acoustic impulse response

In case of exact modelling scenario we set the filter length $M = M_{\mathrm{A}}$ while in under-modelling scenario we halve the exact length, so $M = 512$. In addition,

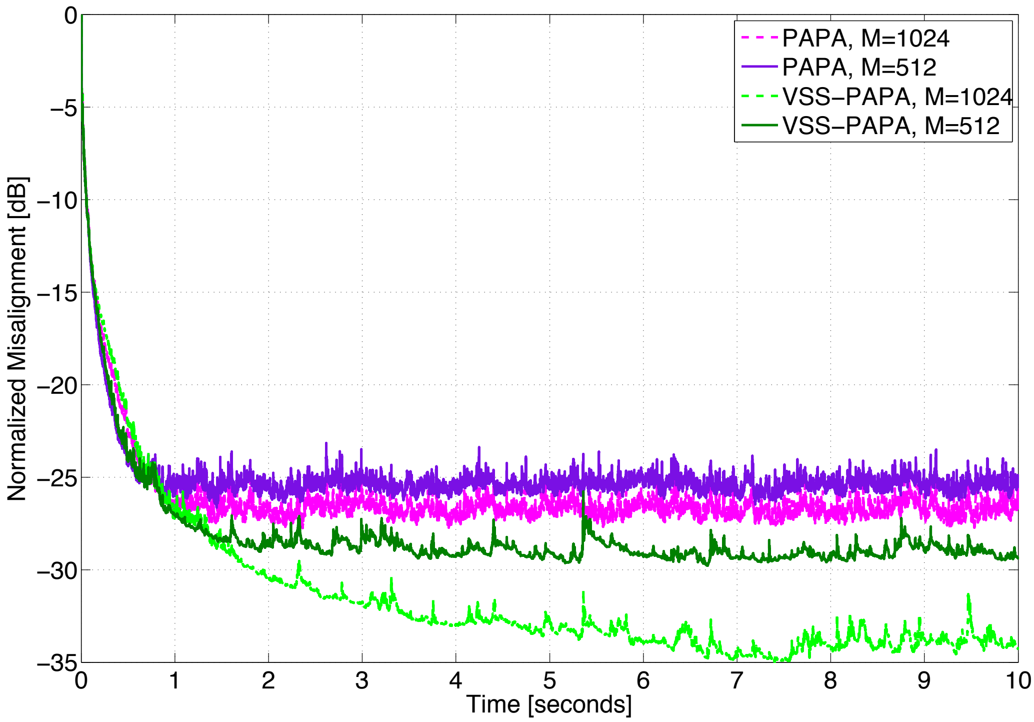

Fig. 6.10: *Misalignment comparison between PAPA and VSS-PAPA algorithms with a white Gaussian noise input. Both algorithms are evaluated in either an exact and an under-modelling scenario.*

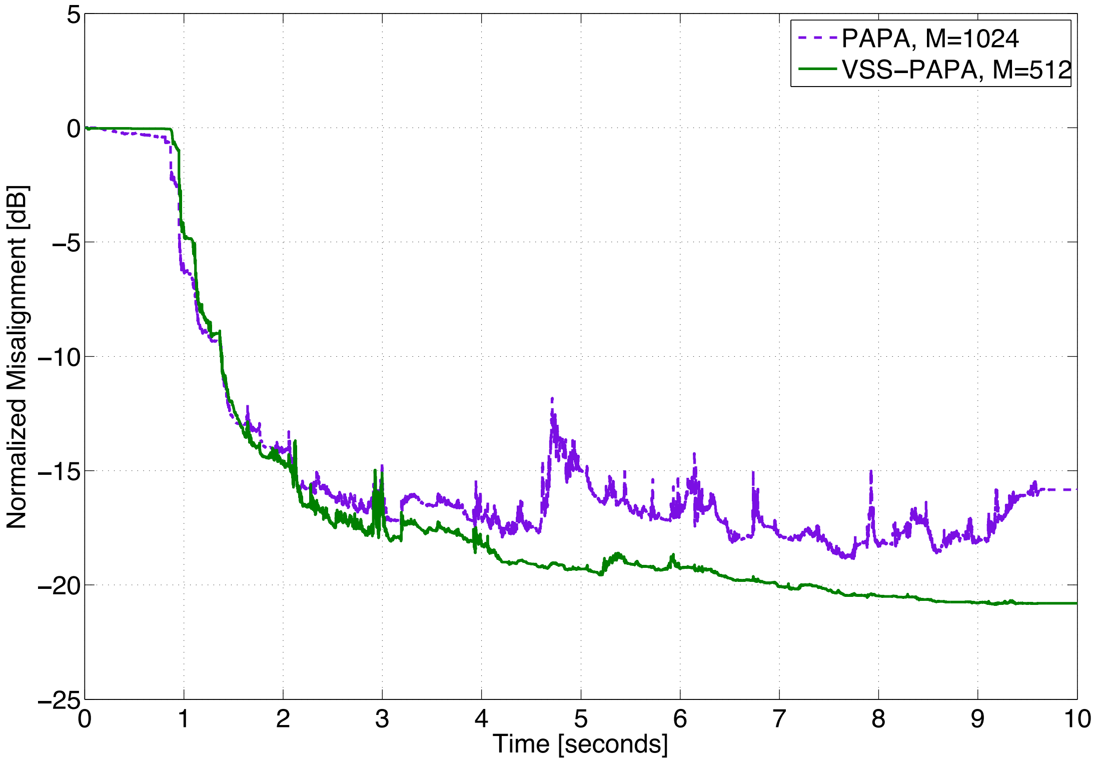

Fig. 6.11: *Misalignment comparison between PAPA and VSS-PAPA algorithms with a female input. The PAPA is evaluated in exact modelling while the VSS-PAPA in under-modelling.*

for the computation of the forgetting factor β in (5.37) we choose $Q = 6$ for white Gaussian noise input and $Q = 20$ for speech input.

In Fig. 6.10 we compare the misalignment performance of PAPA and its VSS version both in exact modelling and under-modelling scenarios using a white Gaussian noise input. It can be notice that even with a strong under-modelling of the AIR the VSS-PAPA achieves better performance compared to PAPAs. In Fig. 6.11 the misalignment comparison is reported in case of speech input using an exact modelling PAPA filter and an under-modelling VSS-PAPA filter; also in this case the VSS-PAPA still outperforms the PAPA. On the other side, not significant improvement is obtained in terms of ERLE, as depicted in Fig. 6.12, however, for an equivalent ERLE the misalignment improvement still represents an advantage since it implies a higher quality of the processed signal in perceptive terms.

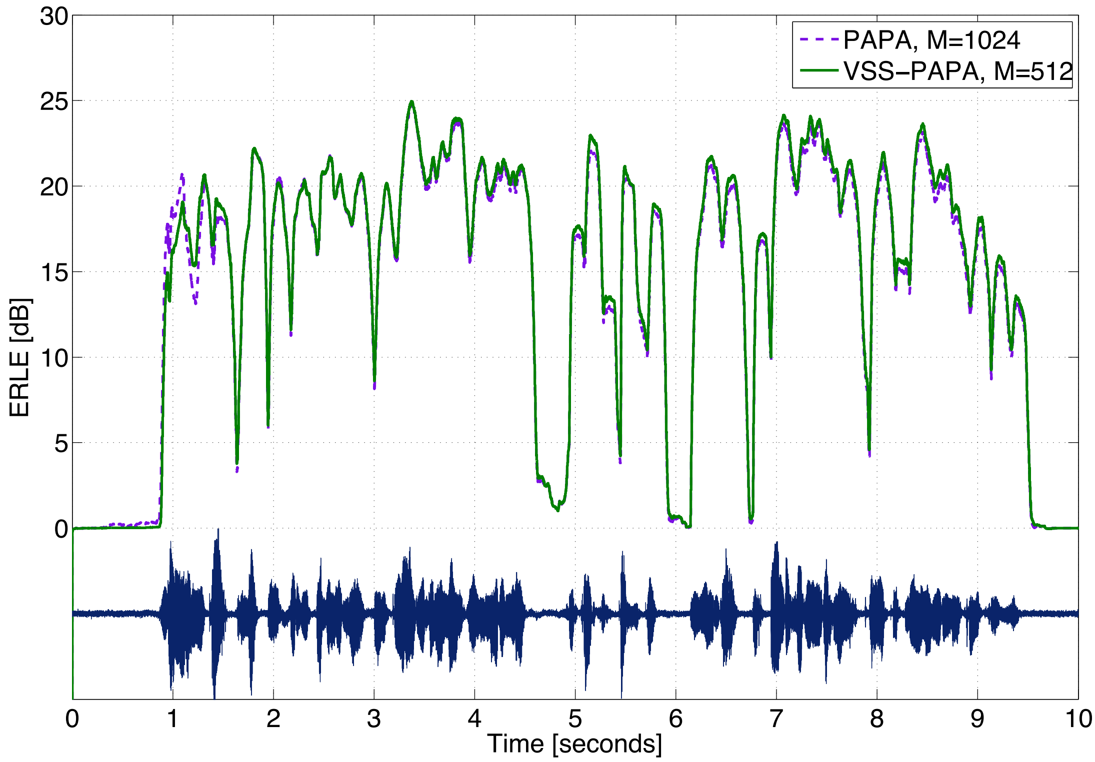

Fig. 6.12: *ERLE comparison between PAPA and VSS-PAPA algorithms with a female input. The PAPA is evaluated in exact modelling while the VSS-PAPA in under-modelling.*

6.3.2 Robustness against double talk

Another situation in which the VSS algorithms result effective is in presence of *double talk*, i.e. when a near-end speech is present and is superimposed over the echo path. In fact in this case it results very difficult to cancel the echo contribution without eating away at near-end speech. The performance of an echo canceller during double talk is an important measurement because near-end speech often causes divergence, especially at high convergence rate. In order to solve this problem a *double talk detector* (DTD) is usually adopted [57], which stops the filter adaptation in presence of double talk in order to preserve the near-end speech. A DTD is a good method to meet the contradictory requirement of low divergence rate and fast convergence in echo cancellation.

DTDs can mostly be classified into energy-based or correlation-based techniques. The most popular representative of energy-based DTDs is the Geigel algorithm [39]. It is based on an observation that the energy of echo is typically much smaller than the energy of far-end speech. Therefore, if the near-end speech is present, the energy of the desired signal increases. The Geigel DTD detects the near-end signals by comparing the magnitude of current far-end sample and the maximum magnitude of the recent past samples of the near-end signals, which means declaring double talk when:

$$|d[n]| = \tau \max\{|x[n]|, \ldots, |x[n-M+1]|\} \tag{6.1}$$

The parameter τ is a threshold usually set to 0.5 based on the assumption of 6 dB hybrid attenuation. Once the double talk is declared, the updates is inhibited for some hangover time in order to reduce the miss of detection.

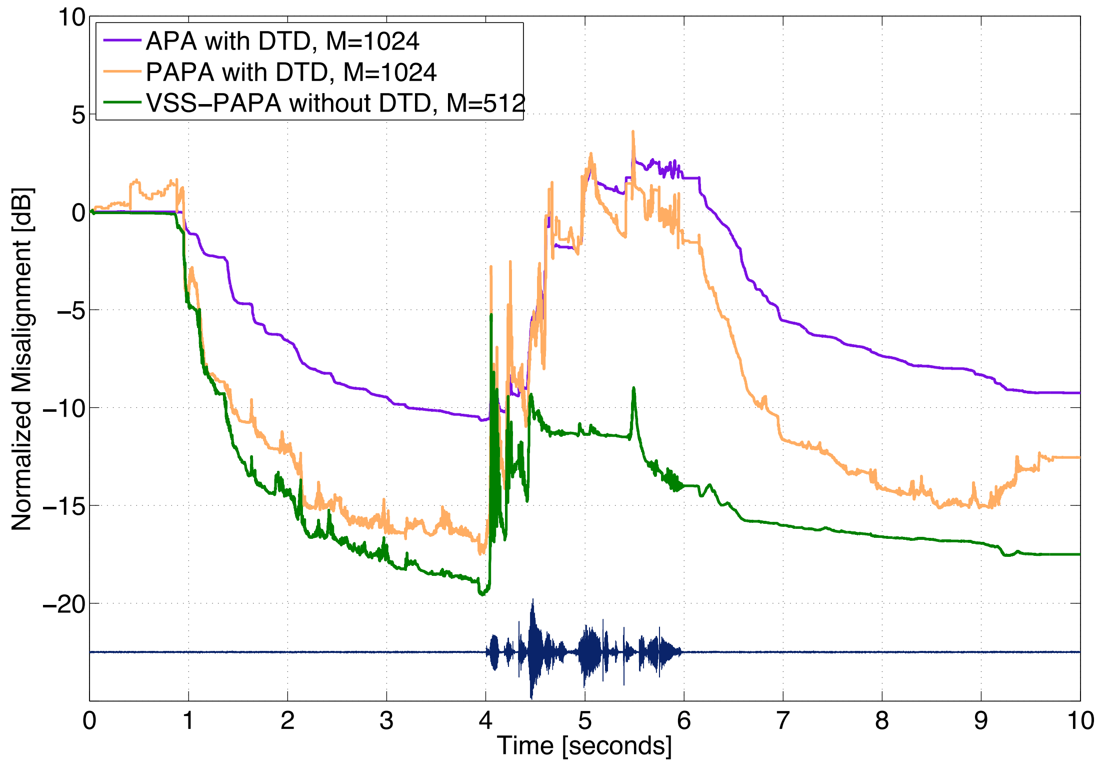

Fig. 6.13: *Misalignment comparison between APA, PAPA and VSS-PAPA algorithms in presence of double talk. APA and PAPA use a Geigel DTD, unlike the VSS-PAPA, which also considers an under-modelling of the AIR. The near-end speech is reported for clearness.*

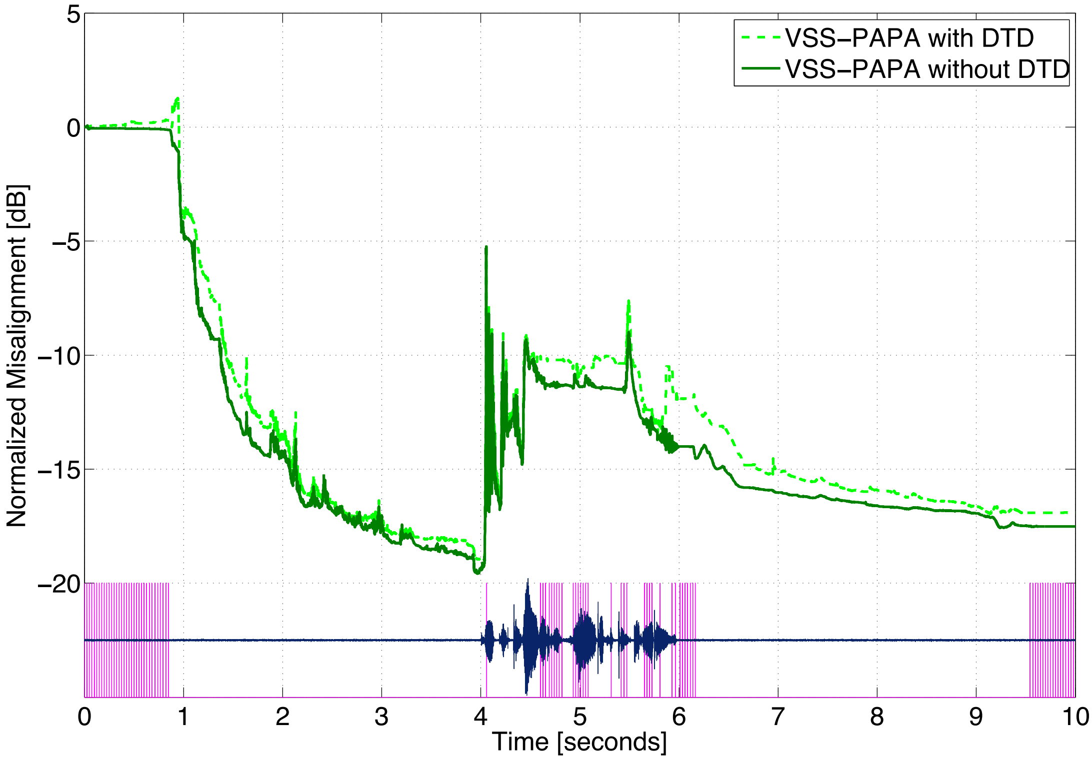

Fig. 6.14: *Misalignment comparison between VSS-PAPA algorithms with and without a DTD in presence of double talk. In both the cases an under-modelling of the AIR is considered. The near-end speech and the double talk detections are reported for clearness.*

However, a DTD is not always a good solution and often it is necessary a strong DTD to preserve the intelligibility of the near-end speech. The strength of the VSS is that it is able to govern the adaptation when a double talk occur, so there is no further need of using any DTD.

Here we consider the same scenario of the previous set of experiments just adding a near-end speech contribution in the time interval $4-6$ seconds in order to simulate a double talk situation. We compare APA, PAPA and VSS-PAPA algorithms in presence of double talk. For APA and PAPA, we use a Geigel DTD with $\tau = 0.5$ and a hangover time equal to 200 samples; on the other side, we use a VSS-PAPA without any DTD and moreover in an under-modelling of the AIR. In Fig. 6.13 it is possible to see that, despite VSS-PAPA is without DTD, it achieves the best misalignment performance compared to other algorithms. In Fig. 6.14, it is possible to verify that a VSS-PAPA without DTD achieves almost the same performance of a VSS-PAPA with DTD, or rather better performance due to the fact that sometimes the DTD may detect a false alarm, so the algorithm stops the adaptation when it should not.