Demodulation as Probabilistic Inference

Richard E. Turner*, *Member, IEEE*, and Maneesh Sahani, *Member, IEEE*

***Abstract*—Demodulation is an ill-posed problem whenever both carrier and envelope signals are broadband and unknown. Here, we approach this problem using the methods of probabilistic inference. The new approach, called Probabilistic Amplitude Demodulation (PAD), is computationally challenging but improves on existing methods in a number of ways. By contrast to previous approaches to demodulation, it satisfies five key desiderata: PAD has soft constraints because it is probabilistic; PAD is able to automatically adjust to the signal because it learns parameters; PAD is user steerable because the solution can be shaped by user-specific prior information; PAD is robust to broad-band noise because this is modelled explicitly; and PAD's solution is self-consistent, empirically satisfying the Carrier Identity property. Furthermore, the probabilistic view naturally encompasses noise and uncertainty, allowing PAD to cope with missing data and return error bars on carrier and envelope estimates. Finally, we show that when PAD is applied to a bandpass filtered signal, the stop-band energy of the inferred carrier is minimal, making PAD well-suited to sub-band demodulation.**

***Index Terms*—Carrier, demodulation, envelope, inference, learning.**

I. Introduction

DEMODULATION is the process by which a signal (y_t) is decomposed into a product of a slowly varying envelope or modulator signal (m_t) and a quickly varying carrier signal (c_t), that is,

$$\mathrm{y}_t = \mathrm{m}_t \mathrm{c}_t. \tag{1}$$

Demodulation was originally developed for radio communications where the carrier is a sinusoid of known frequency, but it has since been applied to a range of audio processing problems including voice coding [1], [2], speech recognition [3], [4], music retrieval [5], speech enhancement [6] and source separation [7], [8], and it is used in hearing devices [6], [9]. Most of these methods employ representations derived from demodulating the sub-bands of the signal. The spectrogram, a widely used representation of signals, can be viewed as such a representation [10] further highlighting the importance of demodulation. Demodulation methods have also been used by researchers to investigate the relative importance of the sub-band envelopes and sub-band carriers (known collectively as the fine-structure) in the perception of sounds [9], [11]–[16]. However, the conclusions that can be drawn from these studies are limited due to several well known problems with demodulation methods [6], [17]–[19].

R. E. Turner is with the Computational and Biological Learning Lab, University of Cambridge, Cambridge, CB2 1PZ, England (e-mail ret26@cam.ac.uk).

M. Sahani is with the Gatsby Computational Neuroscience Unit, University College London, London, WC1N 3AR, England (e-mail maneesh@gatsby.ucl.ac.uk)

Manuscript received July 1, 2010

The central problem for any demodulation algorithm is that, in its most general form, the demodulation problem is ill-posed [20]; any modulator that is non-zero wherever the signal is non-zero can be associated with a valid carrier, and *vice versa*. Thus to achieve repeatable results an algorithm must impose implicit or explicit assumptions about the form of carrier and envelope, often embodied by a set of constraints. For instance, in an amplitude-modulated radio signal the carrier is a sinusoid of known frequency very much higher than the pass-band of the modulator. Imposing this knowledge makes the demodulation problem well posed and straightforward. Unfortunately, in applications involving natural audio signals, there is no clear separation between the carrier and modulator bands and so a more sophisticated approach to designing constraints is required.

Arguably, a general approach to demodulation should impose constraints on the component signals that are soft (that is, violations incur penalties but do not necessarily rule out a candidate decomposition), and that can adapt automatically to the signal, but which are still steerable if required. Both softness and adaptability are needed to handle the variability and potential non-stationarity of the components of the signal. They allow the algorithm to identify suitable bandlimits for the carrier and envelope signals from the measured sound, and to permit temporary or otherwise minor violations of these limits if the resulting solution is better in an overall sense. At the same time, specific knowledge about the properties of the signal generators or desired decomposition may provide partial or approximate information about component properties in some applications. In such cases it would be valuable if this knowledge could be used to steer the outcome of the demodulation algorithm.

At least two further properties seem desirable in the context of natural audio demodulation. The first is robustness to additive noise. Natural signals are often corrupted by broadband noise, and it is clearly desirable for this to have minimal impact on the recovered modulator [21]. Joint demodulation and denoising is essential for many practical applications. A second is that the algorithm be self-consistent. There are many types of consistency [22], but of particular interest is a criterion we call Carrier Identity, which requires that demodulating a recovered carrier yield a constant envelope signal.

Thus, we have proposed five desiderata for a demodulation algorithm in the context of natural signals:

1) soft constraints
2) automatic adjustment to the signal
3) user steerability
4) robustness to noise
5) self-consistency

We argue in section II that existing approaches to demodulation fail to meet these desiderata. Therefore, in the rest of the paper, we develop a new approach called Probabilistic Amplitude Demodulation (PAD) that naturally satisfies them all.

PAD is a new framework which views demodulation as problem of inference and learning. In this context, borrowed from the field of Machine Learning, inference means estimation of variables whose number grow with the number of time-steps. In PAD this is the process of estimating the modulator and carrier from the signal. Learning means estimation of parameters whose number does not depend on the number of time-steps. In PAD this involes estimating properties like the time-scale of the modulator and the modulation depth from the signal. So, whereas inference corresponds to demodulation, learning corresponds to adaptation to the signal.

The starting point of PAD is to articulate a probabilistic forward model (see section III-A) which is a statistical description of the carrier, modulator, and the way in which they combine to form the signal. Bayesian probabilistic calculus is then used to invert the forward model and thereby estimate the carrier and the modulator from the signal (see section III-B). The PAD solution is shaped by the assumptions specified in the forward model, but these constraints are imposed in a soft fashion because they are probabilistic. Furthermore, we show in section III-C that the parameters of the model, like the time-scale of the modulator, can be learned from the signal using techniques like maximum-likelihood estimation. This enables the algorithm to automatically adapt to novel signals. Additionally, any knowledge the user has about these parameters can be incorporated in prior distributions, thereby enabling the algorithm to be steered. In sections IV-A and IV-B we show that PAD is robust to noise. In fact, it is simple to incorporate the noise explicitly in the forward model and, if not known *a priori*, learn its level from the signal. Finally, sections IV-A and IV-B demonstrate that PAD is self consistent in the sense that it approximately satisfies the important Carrier Identity property.

Thus PAD meets all the desiderata we have laid out, but this comes at the price of increased computational cost. PAD uses a range of well-established, but computationally demanding methods for probabilistic inference. For instance, demodulation requires the iterative optimisation of a non-linear cost function. A main focus of this research has been to accelerate the algorithm and currently signals with a sampling rate of 16KHz can be demodulated in real time on a standard laptop.

Although the new approach is computationally challenging it does bring with it several advantages over and above the desiderata mentioned earlier. Unlike many existing demodulation algorithms it serves to make the unavoidable assumptions that determine the solution explicit through the specification of the forward model. This makes PAD easy to understand, critique, and improve. Moreover, PAD can return error-bars on the estimated modulators and carriers. These are especially relevant in signals which are noisy or which contain quickly varying modulators, because there can be considerable uncertainty in the carrier and modulator estimates for these signals. The ability to handle uncertainties enables the range of demodulation tasks to be generalised, for example to signals containing missing regions in which the modulator must be filled-in (see sections IV-A and IV-B).

II. Background

There are many existing demodulation algorithms, but as we argue below, no single algorithm satisfies all of the desiderata introduced in the previous section. We limit our discussion to methods which return positive envelope signals as this is the focus of the paper.

Two classic techniques are the Square and Low-Pass (SLP) method [23] and the Hilbert Envelope (HE) [24]. The SLP method first squares the signal to move modulator energy down to low frequencies where it is then picked off by low-pass filtering. The method is exact when the signal is composed of a high-frequency narrow-band carrier, and a low frequency modulator. When applied to more complex signals, a reasonable modulator can be extracted by judicious choice of the low-pass filter cut-off. However, the recovered carrier is often poor. This is because the envelope often becomes small, or even zero, in regions where the signal is non-zero and this causes the associated carrier to be very large, even unbounded. Thus the method fails desiderata 1, 2, and 5.

The failure of the SLP method to return bounded carrier estimates, and the need to set the low-pass filter, motivate the development of new demodulation method, which is guaranteed to return a bounded carrier and requires no hand-tuning. The HE, given by the magnitude of the analytic signal formed from the measurements, is one such method. Like the SLP method, it is invariant to amplitude scale changes, and a pure tone has constant HE, both of which are useful theoretical properties. However, it still suffers from several problems. Practically, the method performs poorly when the carriers are not single tones. For example, if the signal is a pair of harmonically related sinusoids that undergo slow modulation, the HE will contain a contribution at the fundamental frequency no matter how slowly the true envelope varies. Consequently, the HEs extracted from natural sounds often contain pitch information, even though many applications seek to separate this from modulation content [21]. Similarly, the HE is sensitive to noise in the signal, and therefore not robust. The HE has theoretical problems too. For example, the Hilbert carrier can be discontinuous for continuous signals. Furthermore, the Hilbert carrier is not limited to the same frequency region as the signal—which leads to reconstruction problems when using demodulated sub-bands of a signal [17]. With regard to the desiderata enumerated here, the method fails them all.

An alternative approach to demodulation is to focus on estimating the carrier and then to recover the modulator by division. Coherent approaches to demodulation [25] make the additional assumption that the carrier is a frequency modulated sinusoid, and so demodulation is achieved by estimating the instantaneous frequency of the carrier, for instance using the spectral centre of gravity of the windowed signal. These methods work well when the carrier is well approximated as a single sinusoid (e.g. when operating on a narrow sub-band of the signal), but like the HE method, they fail when the carrier is more complex. Moreover, setting parameters like the window time-scale is critical and the method would benefit from an automatic procedure. Once again, the method fails all of our desiderata.

In this paper, we view demodulation as a Bayesian inference problem. This perspective was first introduced by Turner and Sahani [26] through a simple version of PAD. The main goal of this original paper was to extend PAD to handle cascades of modulators with different time-scales. This work was then generalised to multi-band PAD [27]. In the current paper, we return to consider single-band demodulation in more depth and extend PAD in several new directions. First we consider a more sophisticated and flexible model than that used in the original work [26]. Second, we provide methods for learning all of the free-parameters of the model, which enable the model to automatically adapt to the signal. Third, we present methods for accelerating inference. The utility of these new methods is then demonstrated on synthetic and natural signals, in complete-data, noisy-data and missing-data tasks.

Probabilistic amplitude demodulation proceeds by optimising a non-linear cost function. This is potentially problematic as the optimisation can be slow and there can be multiple (local) optima. Recently, in an elegant paper, Sell and Slaney [21] develop a more computationally efficient demodulation algorithm that optimises a convex cost function [28] and therefore ensures the problem has a unique solution. In section III-D we show that Sell and Slaney's linear-demodulation algorithm can be viewed as a version of PAD. This perspective is important as it reveals the assumptions implicit in their method and it means that the machinery developed in this paper for learning the free-parameters of PAD models can also be applied in the convex case, allowing it to automatically adapt to the signal.

III. Probabilistic Amplitude Demodulation

This section covers the theoretical development of PAD. In order to prevent the main ideas from being obscured by technical detail, we begin this section with a high-level road-map, highlighting the relationship with the desiderata.

The starting point is the forward model [29], which is a description of the process which we assume generated the signal. In the present context, the forward model assumes that, (1) the signal is formed from a product of a modulator and a carrier, (2) the carrier is quickly varying, and (3) the modulator is slowly varying and positive. This information is encoded probabilistically in, (1) the likelihood $p(\mathrm{y}_{1:T}|\mathrm{c}_{1:T}, \mathrm{m}_{1:T}, \theta)$, (2) the prior over the carrier, $p(\mathrm{c}_{1:T}|\theta)$, and (3) the prior over the modulators, $p(\mathrm{m}_{1:T}|\theta)$. Each of these distributions depend on parameters, θ, which control factors like the typical time-scale of the modulator or the frequency content of the carrier. In order to test these modelling assumptions, samples can be drawn from the forward model (see Fig. 1). In general, a balance has to be struck between the accuracy of the modelling assumptions and the tractability of inference.

The forward model specifies the joint probability of the signal, carrier and modulator

$$p(\mathrm{y}_{1:T}, \mathrm{c}_{1:T}, \mathrm{m}_{1:T}|\theta) = p(\mathrm{y}_{1:T}|\mathrm{c}_{1:T}, \mathrm{m}_{1:T}, \theta)p(\mathrm{c}_{1:T}|\theta)p(\mathrm{m}_{1:T}|\theta). \quad (2)$$

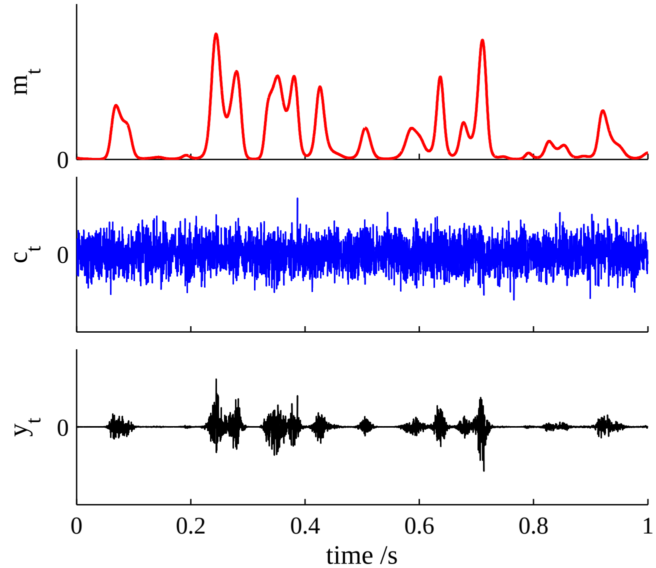

Fig. 1. A sample from the GP-PAD forward model produced using parameter values learned from a natural speech sound. The top panel shows the slowly varying envelopes. The middle row of panels show the quickly varying carriers. The bottom row shows the generated signal which is a modulated version of the carrier.

Inference proceeds by using Bayes' theorem to invert the forward model and form the posterior distribution over the modulators and the carriers, given the data,

$$p(\mathrm{c}_{1:T}, \mathrm{m}_{1:T}|\mathrm{y}_{1:T}, \theta) = \frac{p(\mathrm{y}_{1:T}, \mathrm{c}_{1:T}, \mathrm{m}_{1:T}|\theta)}{p(\mathrm{y}_{1:T}|\theta)} \quad (3)$$

The full solution to PAD is therefore a distribution over possible modulators and carriers, and not a single modulator-carrier pair. This simply reflects the fact that there is not sufficient information to solve ill-posed problems unambiguously. For practical applications the posterior distribution must be summarised, and one approach is to return the most probable modulator and carrier given the signal,

$$\mathrm{c}^*_{1:T},\ \mathrm{m}^*_{1:T} = \underset{\mathrm{c}_{1:T}, \mathrm{m}_{1:T}}{\arg\max}\ p(\mathrm{c}_{1:T}, \mathrm{m}_{1:T}|\mathrm{y}_{1:T}, \theta), \quad (4)$$

together with error-bars which indicate the uncertainty around this best-estimate. Demodulation therefore reduces to optimisation of a cost-function which specifies how the various constraints trade-off with one another in a soft manner (desideratum 1).

The parameters of the model, θ, control how the constraints trade-off, and therefore determine the PAD solution. One way to set the parameters is to choose some general purpose values (e.g. [21]), possibly determined by inspecting samples from the forward model. However, this approach runs into problems when there is a mismatch between the modelling assumptions and the signal. In general we would like the model to have a fairly large number of parameters which automatically adjust to the signal (desideratum 2). Fortunately, there are a number of methods for learning parameters of probabilistic models from the signal. Perhaps the simplest is to use the maximum-likelihood (ML) value of the parameters.

$$\theta^{\mathrm{ML}} = \arg\max_{\theta} \; p(\mathrm{y}_{1:T}|\theta), \tag{5}$$

$$= \arg\max_{\theta} \; \int d\mathrm{c}_{1:T} d\mathrm{m}_{1:T} p(\mathrm{y}_{1:T}, \mathrm{c}_{1:T}, \mathrm{m}_{1:T}|\theta). \tag{6}$$

Unfortunately the integral which this demands is often analytically intractable and so numerical approximation methods have to be used. The art is to find accurate, but fast approximations.

If the user has some prior knowledge of the parameters of the model then an alternative to the ML estimate is the MAP estimate. Here, the user's knowledge is incorporated into the prior over parameters, $p(\theta)$ which shapes the solution (desideratum 3),

$$\theta^{*} = \arg\max_{\theta} \; p(\theta|\mathrm{y}_{1:T}) = \arg\max_{\theta} \; p(\mathrm{y}_{1:T}|\theta)p(\theta). \tag{7}$$

and the ML method is often recovered when the prior over parameters is uniform, $p(\theta) = c$.

The next three sections follow the path described in this section, beginning with a mathematical description of the forward model, then considering inference, and ending with learning.

A. Forward Model

The defining feature of probabilistic forward models for amplitude modulation is that the signal is produced from a product of a slowly varying modulator and a quickly varying carrier (see equation 1). However, as real data is often noisy, the forward model also incorporates additive uncorrelated Gaussian noise. This term improves the robustness of the method. The model will be developed under the assumption that the noise is non-stationary so as to be as general as possible.

The quickly varying carrier is assumed to be white noise. Whilst this is often unrealistic (e.g. for speech where the carrier may contain pitch and formant information), the assumption works surprisingly well in practice because a separation between the time-scales of the carrier and modulator is sufficient to facilitate accurate inference. The slowly varying modulator process is constrained to be positive in this work. It is produced by taking a slowly varying real-valued (positive and negative) process—henceforth called the transformed modulator (x_t)—and passing it through a static positive non-linearity. To avoid a scale degeneracy between the carrier and modulator, the carrier scale (or equivalently its variance) is set to unity.

The complete forward model can be written:

$$p(\mathrm{x}_{1:T}|\mu_{1:T}, \Gamma_{1:T,1:T}) = \mathrm{Norm}(\mathrm{x}_{1:T}; \mu_{1:T}, \Gamma_{1:T,1:T}), \tag{8}$$

$$\mu_t = \mu, \quad \Gamma_{t,t'} = \gamma_{|t-t'|}, \tag{9}$$

$$\mathrm{m}_t = \mathrm{m}(\mathrm{x}_t) = \sigma_{\mathrm{m}} \log(1 + \exp(\mathrm{x}_t)), \tag{10}$$

$$p(\mathrm{c}_t) = \mathrm{Norm}(\mathrm{c}_t; 0, 1), \tag{11}$$

$$p(\mathrm{y}_t|\mathrm{m}_t, \mathrm{c}_t, \sigma^2_{\mathrm{y,t}}) = \mathrm{Norm}(\mathrm{y}_t; \mathrm{m}_t\mathrm{c}_t, \sigma^2_{\mathrm{y,t}}). \tag{12}$$

The slowly varying transformed modulators are produced from a stationary Gaussian process (see [30] for an introduction to the use of Gaussian processes in machine learning) and so this form of PAD is called Gaussian Process PAD (GP-PAD). A standard choice for the transformed modulator covariance function is the squared-exponential kernel,

$$\gamma_{|t-t'|} = \sigma^2_{\mathrm{x}} \exp\left(-\frac{1}{2\tau^2_{\mathrm{eff}}}(t-t')^2\right), \tag{13}$$

where the parameter τ_{eff} defines the time-scale of a typical sample drawn from the Gaussian Process. The transformed modulators are passed through a 'soft threshold-linear' function to produce the modulators—the non-linearity is exponential, and therefore small, for large negative values of x, and linear for large positive values. This modifies the Gaussian marginal distribution of the transformed modulators into a sparse distribution over the modulator, which is often a good match to the modulator histogram derived from natural sounds.

B. Inference

Exact inference is analytically intractable in GP-PAD because of the two non-linearities in the forward model (equations 10 and 12). A simple approximation is to integrate out the carrier and find the most probable setting of the transformed modulator variables given the signal,

$$\mathrm{x}^{*}_{1:T} = \arg\max_{\mathrm{x}_{1:T}} p(\mathrm{x}_{1:T}|\mathrm{y}_{1:T}, \theta), \tag{14}$$

$$= \arg\max_{\mathrm{x}_{1:T}} \log p(\mathrm{y}_{1:T}, \mathrm{x}_{1:T}|\theta) = \arg\max_{\mathrm{x}_{1:T}} \mathcal{L}(\mathrm{x}_{1:T}).$$

This is slightly different from the scheme described earlier, which optimised over both the modulator and the carrier (see equation 4), but experiments indicate that this approach is both faster and more robust to over-fitting.

There is no closed-form solution for the optimisation problem, but a gradient based method can be used to find a local maximum. The objective-function and the gradients of that function can be computed efficiently, by noting that the objective can be split into a component derived from the likelihood and a component from the prior,

$$\mathcal{L}(\mathrm{x}_{1:T}) = \sum_{t=1}^{T} \log p(\mathrm{y}_t|\mathrm{x}_t, \theta) + \log p(\mathrm{x}_{1:T}|\theta). \tag{15}$$

The likelihood component is simple and fast to compute as $p(\mathrm{y}_t|\mathrm{x}_t, \theta) = \mathrm{Norm}(\mathrm{y}_t; 0, \mathrm{m}^2_t + \sigma^2_{\mathrm{y,t}})$. The component from the prior is more challenging as it involves inverting the $T \times T$ covariance matrix of the Gaussian Process which is intractable for time-series of even modest length ($T > 1000$).

One way around this obstacle is to introduce a new set of unobserved variables, $\mathrm{x}_{T+1:T'}$ where $T^{'} > T$, which have a special relationship with the observed variables. More specifically, they are chosen so that the complete set of augmented variables, $\mathrm{x}_{1:T'}$ are circularly correlated (so, for example, x_1 and $\mathrm{x}_{T'}$ are neighbours). This places the augmented latent variables on a ring and the new covariance matrix, $\Gamma_{1:T',1:T'}$, becomes circulant, $\Gamma_{t,t'} = \Gamma_{\mathrm{mod}(t-t',T')}$. This leads to efficient computation using the Fast Fourier Transform (FFT):

$$\mathcal{L}(\mathrm{x}_{1:T'}) = c + \sum_{t=1}^{T} \log \mathrm{m}_t - \frac{1}{2}\sum_{t=1}^{T} \frac{y_t^2}{\mathrm{m}_t^2} - \frac{1}{2T'}\sum_{k=1}^{T^{'}} \frac{|\Delta\tilde{\mathrm{x}}_k|^2}{\tilde{\gamma}_k}.$$

Where $\Delta\tilde{\mathrm{x}}_k$ is the Discrete Fourier Transform (DFT) of the mean shifted transformed-modulators $\Delta\mathrm{x}_t = \mathrm{x}_t - \mu$, and $\tilde{\gamma}_k$ is the DFT of the covariance function, which is the spectrum of the Gaussian Process:

$$\Delta\tilde{\mathrm{x}}_k = \sum_{t=1}^{T^{'}} \mathrm{FT}_{k,t}(\mathrm{x}_t - \mu), \quad \tilde{\gamma}_k = \sum_{t=1}^{T^{'}} \mathrm{FT}_{k,t}\gamma_t, \tag{16}$$

$$\mathrm{FT}_{k,t} = \exp(-2\pi i(k-1)(t-1)/T^{'}). \tag{17}$$

The derivatives can be computed using the expressions above and are omitted for brevity (see [31] for the details). The conjugate gradient method can be used for optimisation [32]. There is some freedom for setting $T^{'}$, but two useful rules of thumb are that it should be a power of two to accelerate the FFT, and it should be larger than T by several time-scales to avoid wrap-around artifacts arising from the fact that the variables at sample $T^{'}$ are neighbours of those at sample 1.

One of the advantages of framing demodulation as an inference problem is that it leads to methods for estimating the uncertainty in the recovered modulators. This section will now describe how to use an approximate version of Laplace's method (itself an approximation) to do this. Laplace's method approximates the posterior distribution over transformed amplitudes by a Gaussian centred at the true posterior mode and with a covariance matrix given by the negative inverse of the Hessian, H, of the log-joint [29],

$$\begin{aligned} p(\mathrm{x}_{1:T'}|\mathrm{y}_{1:T},\theta) &\approx p(\mathrm{y}_{1:T},\mathrm{x}^*_{1:T'}|\theta) \\ &\times \exp\left(\frac{1}{2}(\mathrm{x}_{1:T'} - \mathrm{x}^*_{1:T'})^{\mathsf{T}}\mathrm{H}(\mathrm{x}_{1:T'} - \mathrm{x}^*_{1:T'})\right), \end{aligned} \tag{18}$$

where,

$$\mathrm{H}_{t,t'} = \frac{d^2}{d\mathrm{x}_t d\mathrm{x}_{t'}} \log p(\mathrm{y}_{1:T},\mathrm{x}_{1:T'}|\theta)\Big|_{\mathrm{x}_{1:T'}=\mathrm{x}^*_{1:T'}}. \tag{19}$$

Laplace's approximation thus provides an estimate of the posterior uncertainty ($\Sigma^{\mathrm{post}} = -\mathrm{H}^{-1}$) and it can also be used to perform an approximate integration of the transformed modulators which is useful for parameter learning,

$$p(\mathrm{y}_{1:T}|\theta) = \int d\mathrm{x}_{1:T'}\, p(\mathrm{y}_{1:T},\mathrm{x}_{1:T'}|\theta), \tag{20}$$

$$\approx p(\mathrm{y}_{1:T},\mathrm{x}^*_{1:T'}|\theta)\frac{(2\pi)^{T-1}}{\sqrt{\det(-\mathrm{H})}}. \tag{21}$$

Unfortunately, the Hessian is a $T^{'} \times T^{'}$ matrix and so exact inversion is typically intractable, necessitating a further approximation. Fortunately, the simple structure of the Hessian makes this easy. Specifically, H comprises a diagonal term from the likelihood (D), and a term from the prior, which is the inverse covariance matrix,

$$\mathrm{H}^{-1} = -\Sigma^{\mathrm{post}} = (\mathrm{D} + \Gamma^{-1})^{-1} = \Gamma(\mathrm{D}\Gamma + I)^{-1}, \tag{22}$$

$$= \Gamma^{1/2}(\Gamma^{1/2}\mathrm{D}\Gamma^{1/2} + I)^{-1}\Gamma^{1/2}. \tag{23}$$

This new form is helpful because the difficult inversion is limited to the matrix $A = \Gamma^{1/2}\mathrm{D}\Gamma^{1/2}$ (the other terms being simple to compute exactly). The matrix A inherits the property from the prior covariance Γ that only the low-frequency components have large magnitude. Consequently A can be well approximated by a truncated eigenexpansion, $A \approx \sum_{k=1}^{K_{\mathrm{MAX}}} \lambda_k \mathbf{e}_k \mathbf{e}_k^T$, and the problem reduces to finding an efficient method to compute the top K_{MAX} eigenvectors and eigenvalues of A. Fortunately, the Lanczos algorithm can do just this, requiring only multiplications of A times a vector [33]. These multiplications can themselves be computed rapidly using the FFT. The eigenvalues and vectors can be used to approximate the posterior covariance,

$$\Sigma^{\mathrm{post}} \approx \Gamma^{1/2}\left(I - \sum_{k=1}^{K_{\mathrm{MAX}}} \frac{\lambda_k}{\lambda_k + 1}\mathbf{e}_k\mathbf{e}_k^{\mathsf{T}}\right)\Gamma^{1/2}, \tag{24}$$

This expression has an instructive interpretation; in order to compute the approximate posterior covariance (the posterior uncertainties), begin with the marginal covariance of the prior (the prior uncertainties) and subtract uncertainty from it as more eigenvalues are considered. In practice the posterior variances are returned rather than the full posterior covariance because of memory constraints.

This completes the description of inference, that is estimation of the modulator and the modulator uncertainty. In the next section we describe how to learn the parameters in the model.

C. Learning

The parameters of the GPPAD model embody the assumptions being made about the statistics of the signal, like the typical time-scale of the modulation or the typical sparsity. When faced with a new signal, it is important to set the parameters so that they are in approximate agreement with the statistics of the signal otherwise there is a mismatch which can cause PAD to return poor modulator estimates (see section III). However, as the model is non-linear, it is often difficult to set the parameters by hand. Fortunately, machine-learning provides several methods for automatically determining the parameters.

Any candidate parameter learning scheme must retain uncertainty information in the latent variables in order to avoid over-fitting [31]. One approach that meets this criteria is to select the parameters which maximise Laplace's approximation to the likelihood of the parameters (equation 21). The modulator time-scale (τ_{eff}) will be learned in this way. However, it is computationally costly to evaluate this objective and so a cheaper alternative is used to learn the other parameters (σ_{x}^2, σ_{m}^2 and μ). The approach is motivated by the observation that the marginal distribution of the signal,

$$p(\mathrm{y}|\sigma_{\mathrm{m}}^2,\mu,\sigma_{\mathrm{x}}^2) = \int d\mathrm{x}\; p(\mathrm{y},\mathrm{x}|\sigma_{\mathrm{m}}^2,\mu,\sigma_{\mathrm{x}}^2), \tag{25}$$

$$= \int d\mathrm{x}\; p(\mathrm{y}|\mathrm{x},\sigma_{\mathrm{m}}^2)p(\mathrm{x}|\mu,\sigma_{\mathrm{x}}^2), \tag{26}$$

does not depend on the temporal dynamics of the model and so it is simple to evaluate numerically (e.g. by gridding up the space that has significant mass under the prior). This leads to a two stage scheme in which the marginal distribution of the signal is first used to learn σ_{x}^2, σ_{m}^2 and μ, followed by a second stage in which Laplace's approximation to the likelihood is used to learn the time-scale of the modulation. Prior information that the user possesses about the parameters can be incorporated into both of these stages using equation 7. The first stage of learning typically takes a few seconds on a standard laptop, independent of the signal duration. The second stage of learning typically takes a few minutes for a second of sound at 16000Hz.

D. Relationship to existing methods

Before demonstrating PAD practically, we note several connections to existing demodulation methods. One key observation is that, for a noiseless signal ($\sigma^2_{\mathrm{y,t}} = 0$), the GPPAD assumptions imply that the square of the signal is an unbiased estimator for the square of the modulator at that time point, $\langle \mathrm{y}_t^2 | \mathrm{m}_t \rangle = \mathrm{m}_t^2$. In principle the square of the signal can therefore be used to estimate the modulator, but it performs poorly because it has a very large variance ($2\mathrm{m}_t^4$). However, this estimator can be converted into a more useful one by leveraging the slowness of the modulator and averaging the squared signal over a local region. This reduces the variance of the estimator at the cost of introducing some bias because the true modulator will vary over the local region. Importantly, this scheme is precisely the SLP method described earlier. From this new perspective, choosing the low-pass filter cut-off amounts to a bias-variance trade-off. Motivated by this connection, the SLP method is used as an intialisation for PAD. We reiterate the fact that, although the SLP method often provides a reasonable estimate of the modulator (in a squared-error sense), the corresponding estimate for the carrier is often extremely inaccurate and so fine-tuning of the SLP solution is essential for many applications.

In a recent paper, Sell and Slaney introduce an elegant approach to demodulation [21] that defines the modulator as the solution to a convex optimisation problem which can be written,

$$\text{minimise } ||W(f)\mathcal{F}(\mathrm{m}(t))||_2^2 \text{ subject to } \mathrm{m}(t) \geq |\mathrm{y}(t)|. \quad (27)$$

Here, $\mathcal{F}$ is the Fourier transform and $W(f)$ is a window function that penalises high-frequency energy in the modulator. The basic idea is that the constraints ensure the modulator is greater than the absolute value of the signal at each time-point, and the cost-function ensures the solution goes near to the rectified signal, in a slowly varying and smooth manner. This sensible scheme was motivated heuristically, but here we show that it can also be derived from a probabilistic model. Consider a forward model for amplitude demodulation in which a positive modulator ($\mathrm{m}_t \geq 0$) is drawn from a truncated multivariate Gaussian and where the carrier is drawn from a uniform distribution,

$$p(\mathrm{m}_{1:T}|\Sigma_{1:T,1:T}) = \frac{1}{Z}\exp\left(-\frac{1}{2}\mathrm{m}_{1:T}^{\intercal}\Sigma_{1:T,1:T}^{-1}\mathrm{m}_{1:T}\right), \quad (28)$$

$$p(\mathrm{c}_t) = \mathrm{Uniform}(\mathrm{c}_t; -1, 1), \quad (29)$$

$$\mathrm{y}_t = \mathrm{m}_t\mathrm{c}_t. \quad (30)$$

The fact that the carriers in this model are bounded between $-1 \leq \mathrm{c}_t \leq 1$ can be motivated from a sinusoidal model, $\mathrm{c}_t = \sin(\phi_t)$.

The prior over the carriers enforces the constraint that $|\mathrm{c}_t| \leq 1$. The likelihood enforces the constraint that, $\mathrm{c}_t = \mathrm{y}_t/\mathrm{m}_t$. Both of these constraints are only satisfied when the modulator is greater than or equal to the data magnitude, $\mathrm{m}_t \geq |\mathrm{y}_t|$. The posterior distribution over modulators is therefore another truncated Gaussian where the constraints define the new truncation points. The MAP modulator is therefore given by,

$$\mathrm{m}^*_{1:T} = \underset{\mathrm{m}_{1:T}}{\arg\min}\, \mathrm{C}(\mathrm{m}_{1:T}) \quad \text{such that} \quad \mathrm{m}_t \geq |\mathrm{y}_t|. \quad (31)$$

where the cost function is the negative of the prior probability of the modulators,

$$\mathrm{C}(\mathrm{m}_{1:T}) = \frac{1}{2}\mathrm{m}_{1:T}^{\intercal}\Sigma_{1:T,1:T}^{-1}\mathrm{m}_{1:T} \approx \frac{1}{2}\sum_{k=1}^{T}\frac{|\tilde{\mathrm{m}}_k|^2}{\tilde{\gamma}_k} \quad (32)$$

Therefore, the MAP modulators are found using a quadratic program, which is a simple convex cost function [28]. This is the same cost function used in Sell and Slaney's 'linear' demodulation algorithm and reveals the connection between their approach and probabilistic models [21]. This connection is important, for example, as it gives rise to methods for learning the free parameters in the model, like their spectral weighting function, $W(f)$, which is equivalent to $\tilde{\gamma}_k^{-1/2}$.

One of the contributions of previous research on demodulation has been to catalogue desirable properties that an ideal demodulation algorithm should have. In the probabilistic approach similar properties arise naturally when the rules of probability are used to invert the generative model. For instance, the carrier and the modulator recovered by PAD from a bounded signal will also be bounded [20] because the prior probability of an unbounded carrier or modulator is 0. Similarly, the modulator will be smooth [24] because a realisation from a GP prior with a squared exponential kernel is (almost surely) analytic [30]. PAD is also covariant with respect to scale changes in the input signal (a generalisation of [24]) because the maximum likelihood setting of the modulator variance rescales to compensate for any change in the signal's scale.

IV. Results

This section applies PAD to various signals, demonstrating that it is in compliance with the remaining desiderata and comparing it to methods which fail them, like the SLP and HE methods. One of the main challenges posed by the evaluation of demodulation algorithms is that ground truth is unknown for natural signals. This means that a quantitative comparison of different schemes must take an indirect approach. We present several different comparisons of this sort. The first approach uses synthetic signals, for which the ground truth carriers and modulators are known (see section IV-A). The results suggest that PAD is more flexible than other methods. In particular, PAD performs well even when the modulator and carrier bands overlap, which is typically the case for natural signals, and also when the signal is stochastic. Although these results are suggestive, they are not conclusive because it is unclear which synthetic signal class is a sensible approximation to natural sounds. In section IV-B PAD is applied to speech and the estimated carriers and modulators are shown to be qualitatively superior to those recovered by other methods. Various measures of consistency can be used to evaluate the solutions, such as demodulating the carrier and testing whether the Carrier Identity property holds. Another consistency measure, which relates to robustness, is to estimate the modulators from a noisy signal and to determine how close they are to those estimated from the clean signal. Furthermore, modulators estimated in missing-data regions should also be close to those estimated from the complete signal. These consistency tests are important criteria that a demodulation algorithm should meet, but it is important to note that they are not sufficient. For example, an algorithm that returns a constant modulator, independent of the signal, would not be a good demodulation algorithm and yet it would pass the tests mentioned above.

In the last part of this section, PAD is applied to the subbands of a signal showing that, in contrast to many existing methods, the carrier remains reasonably band-limited. This is critical for reconstruction [18].

A. Synthetic signals

The experiments described in this section test PAD on synthetic signals in three different settings; noise-free data, noisy-data and missing-data. In all of the experiments, the synthetic signal comprised a modulated carrier, possibly combined with additive Gaussian noise, $y_t = \mathrm{m}_t \mathrm{c}_t + \sigma_{\mathrm{y}_t} \epsilon_t$. Three different carriers were used

1) A sinusoidal carrier, $\mathrm{c}_t^{(1)} = \sin(2\pi f^{(c)} t)$ where $f^{(c)} = 100.7$Hz.
2) A harmonic carrier, $\mathrm{c}_t^{(2)} = \sin(2\pi f_1^{(c)} t) + \sin(2\pi f_2^{(c)} t)$ where $f_1^{(c)} = 100.7$Hz and $f_2^{(c)} = 201.4$Hz
3) A white noise carrier, $\mathrm{c}_t^{(3)} \sim \mathrm{Norm}(0, 1)$

Generally speaking, the properties of the carrier were found to have a more substantial effect on the performance of the demodulation algorithms than the properties of the envelope. For this reason the same envelope was used throughout all of the experiments; an exponentiated sum of three sinusoids,

$$\mathrm{m}_t = \exp\left(\sum_{k=1}^{3}\left(\alpha_k^{(1)} \sin(2\pi f_k^{(m)} t) + \alpha_k^{(2)} \cos(2\pi f_k^{(m)} t)\right)\right).$$

The coefficients of the sinusoids were drawn from a unit variance Gaussian $\alpha_k^{(i)} \sim \mathrm{Norm}(0, 1)$ and the frequencies of the sinusoids uniformly sampled between $0 - 2$Hz, $f_k^{(m)} \sim \mathrm{Uniform}(0, 2)$. The results reported below are robust to the seed of the pseudo random number generator. A sample rate of 2000Hz was used for the experiments.

For reference, PAD is compared to the HE and SLP methods below. One of the metrics used for comparison is the Signal to Noise Ratio (SNR) between the true envelopes (m_t) and the estimated envelopes ($\hat{\mathrm{m}}_t$),

$$\mathrm{SNR_m} = 10 \log_{10} \sum_{t=1}^{T} \mathrm{m}_t^2 - 10 \log_{10} \sum_{t=1}^{T} (\mathrm{m}_t - \hat{\mathrm{m}}_t)^2 \quad (33)$$

A SNR can be defined analogously for the carriers.

Whilst the HE has no free parameters, and can therefore be applied directly to any signal, the SLP method has free parameters that must be chosen, like the low-pass filter cut-off. We choose to use a filter with a logistic shape, $1/(1+exp((f - f_{\text{cut-off}})/f_{\text{width}})$ and set the cut-off ($f_{\text{cut-off}}$) and width (f_{width}) which return the largest SNR for the estimated envelopes. This is an upper-bound on the performance of the SLP method using the logistic filter shape because ground-truth is normally unknown and so the parameters cannot usually be optimised in this way. One further correction is made to the usual SLP algorithm to address the fact that it can return an estimate for the square-envelope that is negative. This can occur because the low-pass filtering of the squared signal is not guaranteed to return a positive signal. To prevent such cases, we modify the method slightly, thresholding such that square-envelopes $\mathrm{m}_t^2 < 10^{-4}$ are set to 10^{-4}.

In order to make the tests as demanding as possible, all of the parameters in PAD were learned from the signal using maximum-likelihood i.e. no prior knowledge was used. For the noisy signals, the noise level was assumed to be known before hand and the parameters were set to those learned from the clean signal.

Some typical results for the three algorithms demodulating three different clean test signals are shown in figure 2 and the results on noisy signals are summarised in figure 3.

The HE is the best estimator when the carrier is a pure tone, although all three methods typically have high SNRs (~ 40) for signals of this type. However, when the carrier is more complex, the HE often contains components that are faster than desired. For example, when carriers contain harmonics (signal 2) the HE contains a component at the fundamental, and when the carriers are stochastic (signal 3) the HE becomes very noisy. For similar reasons, the HE degrades quickly as more noise is added to the signal.

The SLP envelope is more robust than the HE. It is often accurate (as measured by the SNR of the estimated envelopes) so long as it is known where to place the filter cut-off. For this reason, in normal applications it tends to perform well when the carrier and modulator bands are well separated, but less so when there is overlap. Here the threshold is determined automatically, and so this is less of an issue. However, whilst the estimate of the envelope is often accurate in a squared-error sense, the method often returns poor estimates for the carriers (e.g. see signal 3). This happens because in regions of low energy, the SLP envelope can become too small and this results in carriers which are very large. In short, there is no constraint built into the method to ensure the carrier remains well behaved.

PAD out-performs the SLP method in every condition and the HE in every condition bar the simplest (a pure tone carrier in noise free conditions), both in terms of the SNR of the envelopes and the SNR of the carriers (data not shown). PAD approximately satisfies the Carrier Identity test for all of the synthetic signals, whilst the other methods break down catastrophically (as shown in figure 2). The conclusion is that PAD is more robust to changes in the signal class and to additive Gaussian noise.

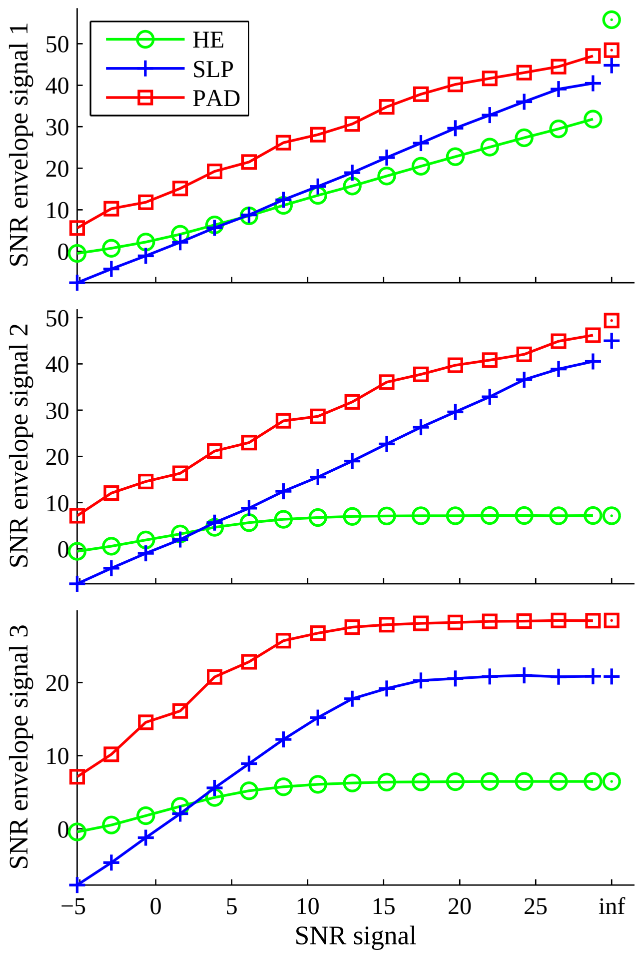

Fig. 3. Demodulating noisy synthetic signals. Noise was added to the signals shown in figure 2 and the envelopes estimated using the three methods. The panels show the SNR of the estimated envelopes as a function of the SNR of the signal. The three methods are; HE (circles), SLP (crosses), PAD (squares). PAD out-performs the other methods on the noisy data by about 3-12dB. The rightmost markers show the performance on the clean signal for reference.

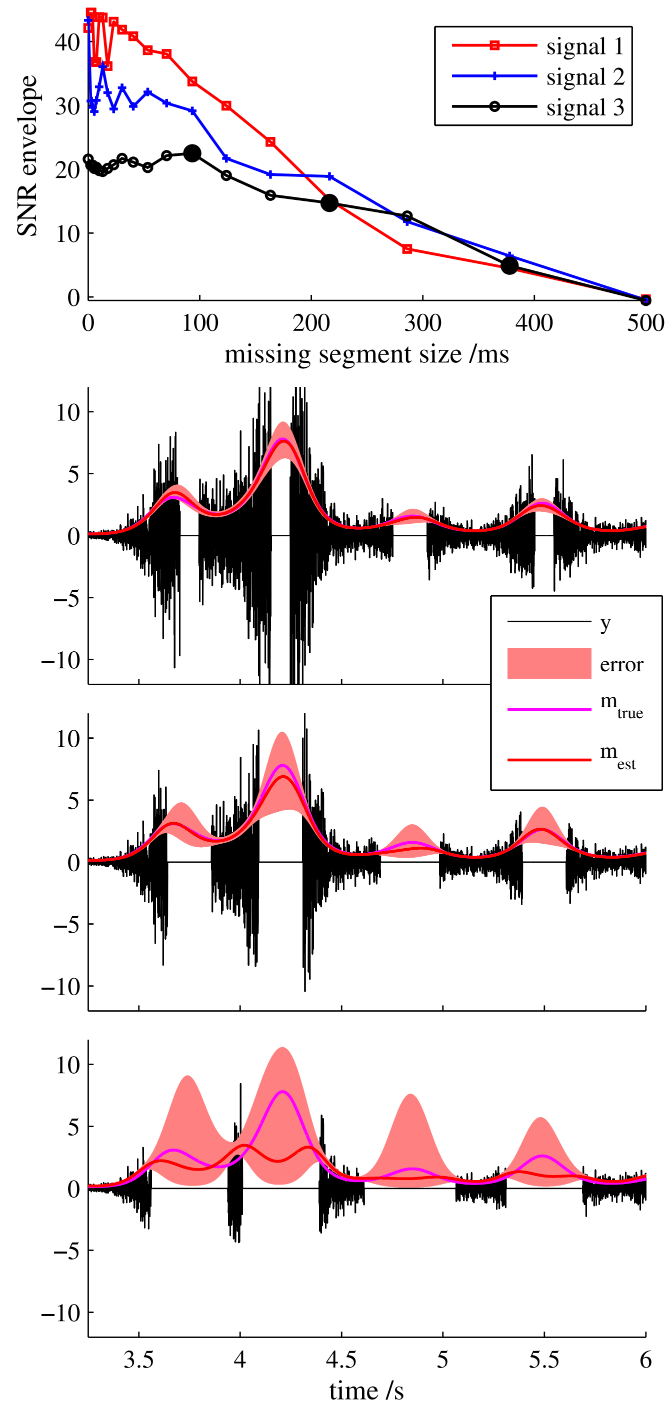

Fig. 4. Demodulating synthetic signals with missing data. Short sections of the three synthetic signals were removed from randomly chosen locations. PAD was used to fill in the missing regions and the SNR of the estimated envelopes in the missing regions was plotted as a function of gap size (top panel) for the three signals, signal 1 (squares), signal 2 (crosses) and signal 3 (circles). The larger filled black circles correspond to the examples plotted in the lower panels for signal 3. The lower panels show a short section of the signal (thin black line), the true envelopes and the estimated envelopes (thick lines) and the error-bars at three standard deviations (shaded region).

Finally, figure 4 demonstrates that PAD can be used to accurately fill in the envelopes in missing regions of the synthetic signals. Practically, the inference proceeds as for complete data, but the variance of the observation noise is set to infinity in the missing region ($\sigma^2_{\mathrm{y,t}} = \infty$). The estimated envelopes are very accurate for small gap sizes, but deteriorate for gaps comparable to the time-scale of the envelope. The uncertainty in the estimated envelopes grows correspondingly with the gap size. Importantly, the true envelope tends to remain within the error-bars at all gap sizes. The conclusion is that PAD can accurately estimate both the envelopes and the uncertainty in regions of missing data.

B. Speech signals

This section applies PAD to a speech sound. Noise-free data, noisy-data and missing-data settings are considered. There is an important difference between the synthetic signals considered in the previous section and natural sounds like speech considered in this section. Whereas the synthetic sounds contained a single time-scale of modulation, natural sounds

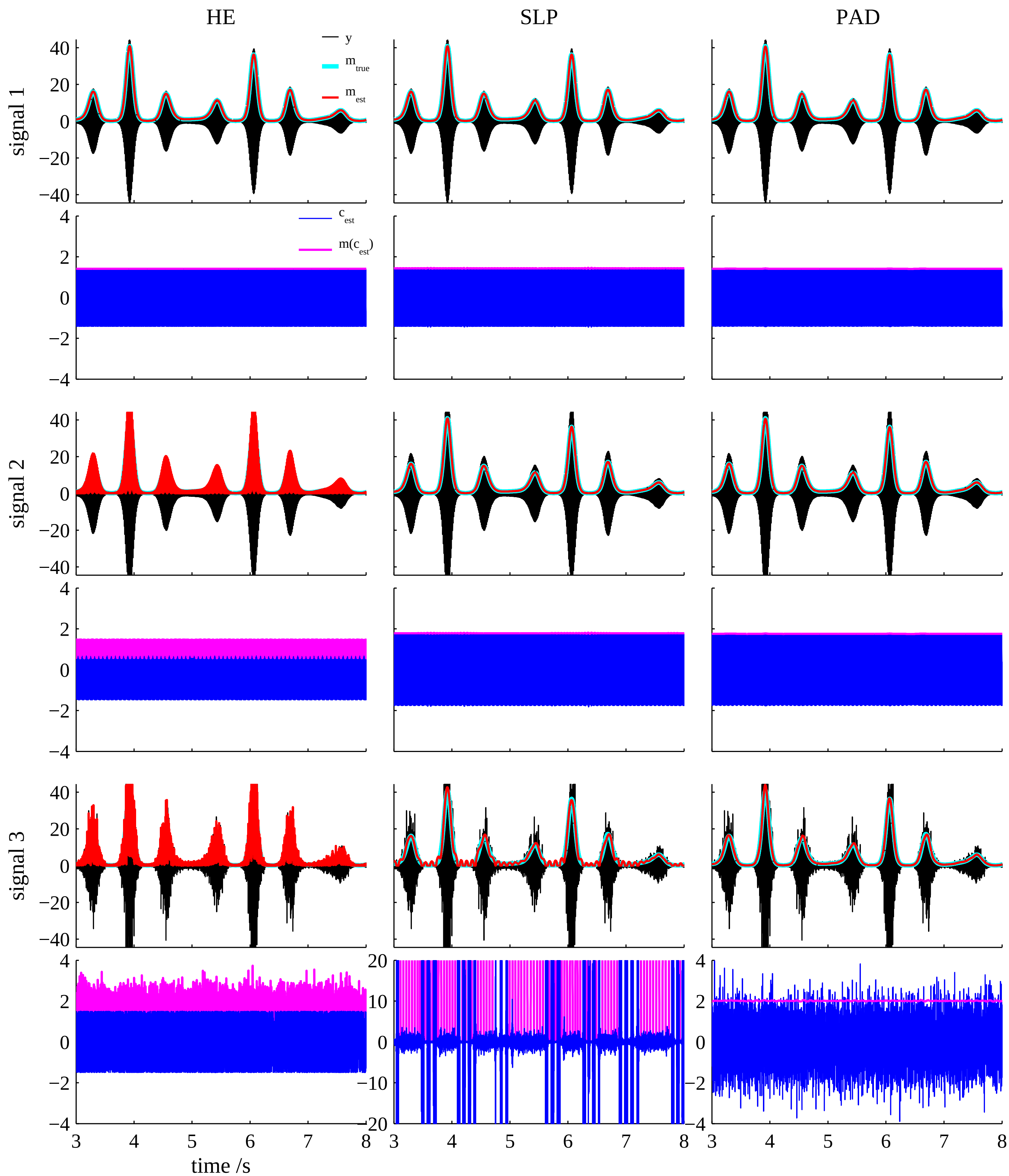

Fig. 2. Demodulating synthetic signals. Each pair of rows shows a different signal (black) decomposed into a modulator (thick line) and a carrier (lower panel) using three different methods; the Hilbert Envelope (column 1), the ALP method (column 2) and PAD (column 3). The true modulator is also shown for reference. The result of demodulating the carriers (an empirical test of the Carrier Identity property) is also shown.

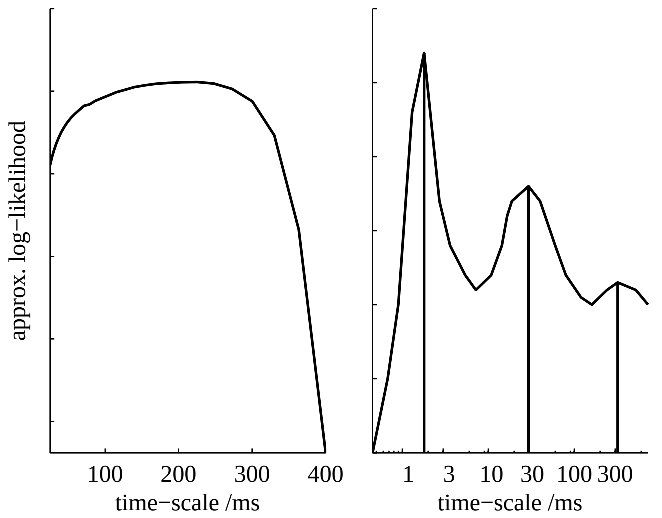

Fig. 5. Approximate likelihoods of the time-scale parameter, τ_{eff}. The left panel shows the log-likelihood for the third synthetic signal which is unimodal. The right panel shows the log-likelihood for the speech signal which has three modes originating from the pitch ($\tau_{\text{eff}}^{-1} = 550$Hz), sylable ($\tau_{\text{eff}}^{-1} = 34$Hz) and sentence ($\tau_{\text{eff}}^{-1} = 3$Hz) structures in speech. These frequencies are higher than normally associated with these structures because each cycle corresponds to two or three time-scales.

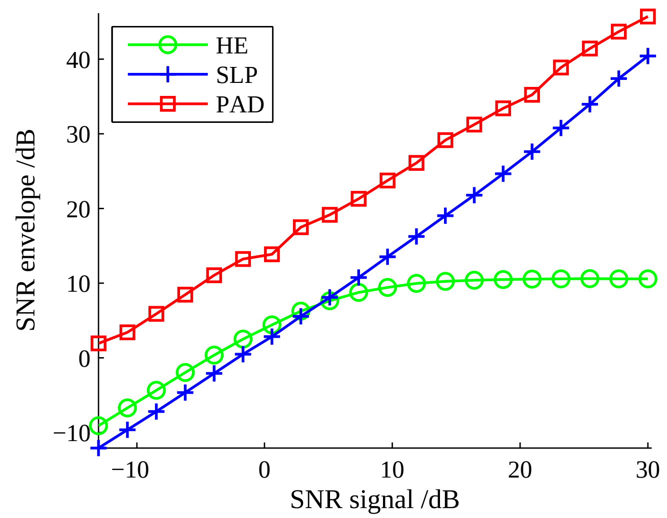

Fig. 8. Demodulating a noisy speech signal. Noise was added to the speech signal shown in figure 6 and the envelopes were estimated using the three methods. The panels show the SNR of the estimated envelopes as a function of the SNR of the signal. The three methods are; HE (circles), SLP (crosses), PAD (squares). PAD out-performs the other methods on the noisy data by about 5-11dB.

often contain modulation at multiple time-scales (see figure 5). Fortunately, PAD can be used to automatically diagnose when this is the case and to select between the various solutions. The key quantity in this process is the (approximate) likelihood of the time-scale, $p(\mathrm{y}|\tau_{\text{eff}})$. This is found to have a single peak for the synthetic sounds considered in the last section indicating a single best solution. However, this quantity has three peaks for the speech sound considered here (see figure 5). Each of these peaks corresponds to modulation arising from a different physical process; the glottal pulse periods, the syllables and the sentences of speech. By appropriately choosing the prior (see equation 7), the user is able to select between these different solutions. For example, a prior which favours the syllable time-scales is used in figure 6. The speech is demodulated effectively and the Carrier Identity property holds to a close approximation. For comparison, when the parameters of the SLP method are chosen to demodulate at this time-scale a reasonable looking modulator can be returned, but the carriers are typically ill-behaved. Similarly, when a prior is used which favours the sentence time-scales, PAD again demodulates the sound fairly effectively (see figure 7), especially in comparison with the result from SLP. Finally, when a prior is used which favours the glottal pulse time-scales, PAD recovers a modulator which resembles the solution provided by the HE. The HE method demodulates voiced phonemes at the time-scale of the vocal fold oscillations because this causes harmonic structure, but in unvoiced sections it becomes noisy.

Next we consider the performance of PAD on signals which are noisy and contain missing data. As ground truth is unknown we compare the envelope estimates derived from the noisy signals to those derived from the clean signals. This is an important consistency test. Figure 8 indicates that the solution from PAD degrades less quickly in the presence of noise than that from the SLP or HE methods. Figure 9 shows that PAD can reliably estimate the envelope of speech signals in missing regions up to 20ms long.

C. Sub-band demodulation

One of the most popular applications of demodulation is to the sub-bands of the signal (see section I). However, it is known that the carriers derived from band-limited signals using many demodulation methods, such as the SLP and HE methods, are not guaranteed to be spectrally limited to the pass-band of the filter (see figure 10 for one example). This is important because it can lead to artifacts when reconstructing signals, for example, by recombining filtered versions of the sub-band modulators with the original carriers [17]. Figure 10 demonstrates that the carriers derived from PAD can be substantially more band-limited than those derived from the HE or SLP methods. However, as there is no constraint on the carrier frequency content, carrier signal energy can still be found outside of the filter. In principle, PAD could be extended to add such a constraint to the spectral content of the carrier, which may improve performance.

An alternative approach to sub-band demodulation, that suggests a different extension to PAD, is that of coherent demodulation [25]. Like the HE, this method assumes the carrier is a single frequency modulated sinusoid, but it is constrained to be limited to the pass-band of the filter. The method performs well when the filters are narrow, but it can perform poorly for broad filters that contain harmonic or noisy carriers which violate the assumptions of the algorithm. Furthermore, there is no positivity constraint on the envelope in this approach and so it is not directly comparable to the approach taken here. In fact, the envelope is completely unconstrained in coherent demodulation. It is interesting to contrast this to PAD where it is the carrier which is unconstrained

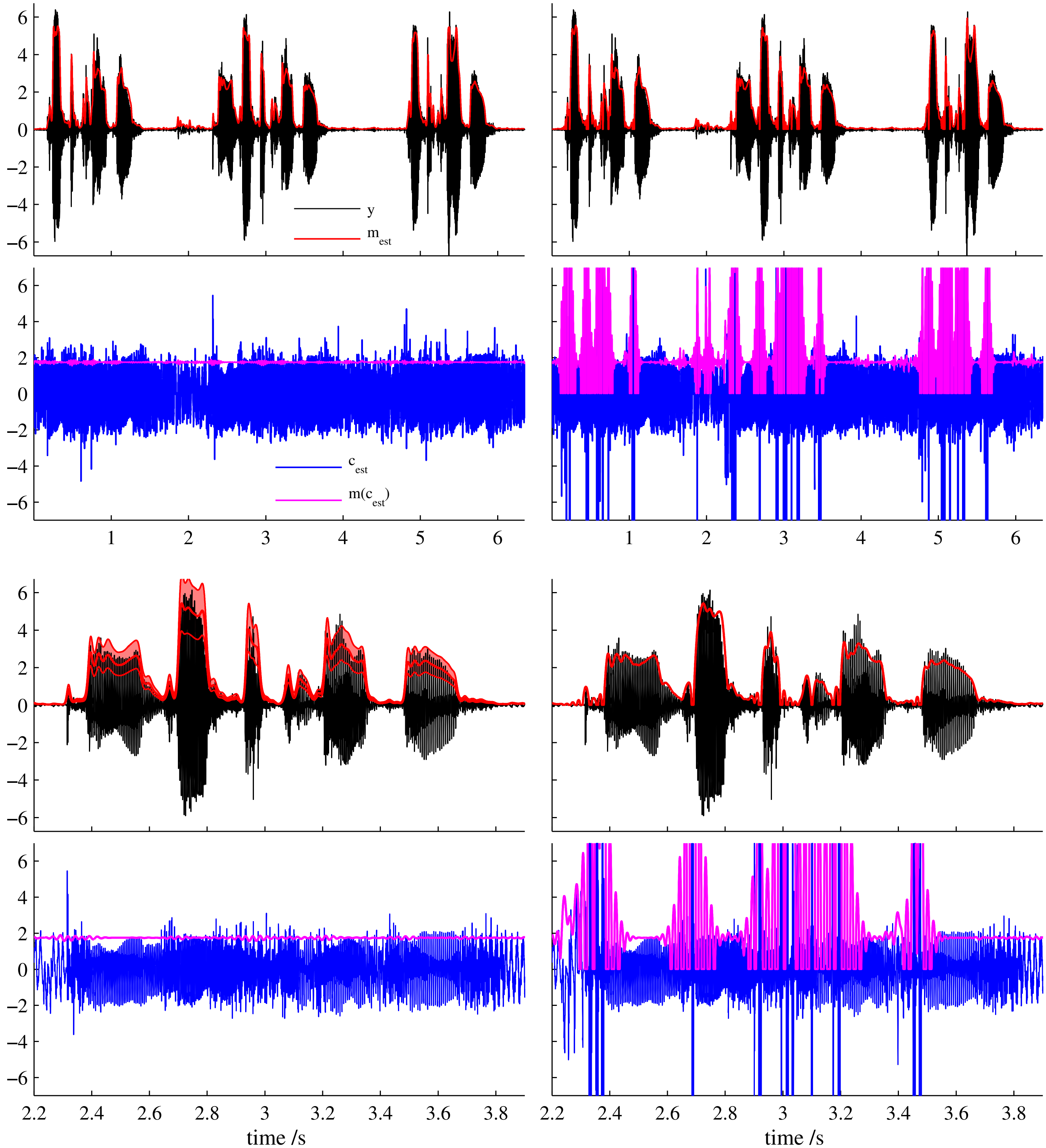

Fig. 6. Demodulating the syllables of speech. The top pair of rows shows a speech signal consisting of three spoken sentences (thin black line, upper panel) decomposed into a modulator (thick positive line, upper panel) and a carrier (thin line, lower panel). The left hand column shows PAD and the right hand column the ALP method. The time-scale used for PAD was learned from the data using a flat prior, and the ALP cut-off was set to give similar results to PAD. The result of demodulating the carriers (an empirical test of the Carrier Identity property) is also shown (thick positive line, lower panel). The bottom pair of panels is a close up of the middle sentence. The error-bars on the PAD envelopes are also indicated by the shaded region.

(beyond having zero mean and unit variance). It appears that a method which imposes constraints on both the carrier and envelope would combine the benefits of both approaches. For instance, one approach in this direction would be to extend PAD so that the carrier is modelled as a frequency modulated sinusoid, and to place a prior over the frequency modulation,

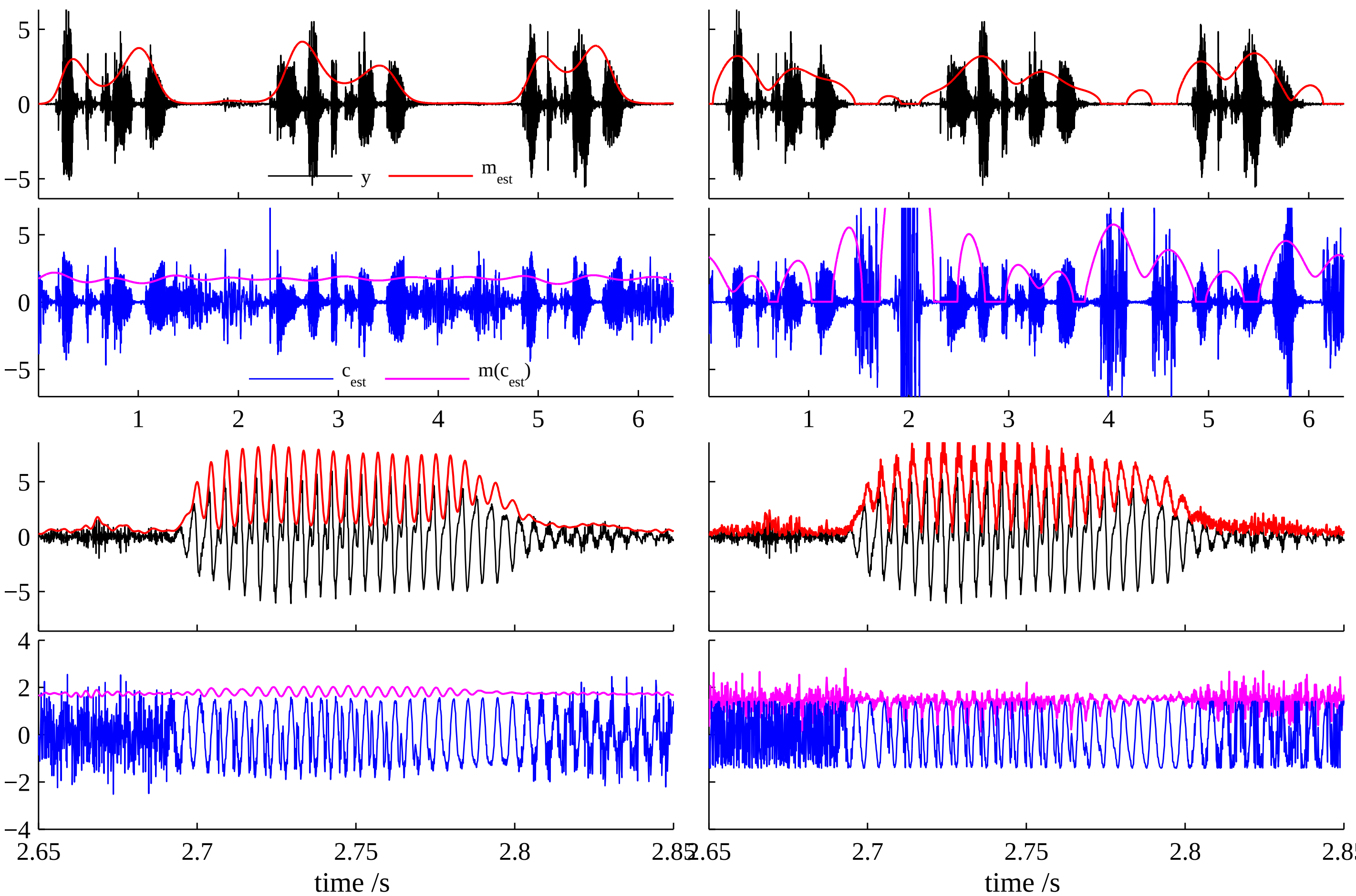

Fig. 7. Demodulating speech at sentence time-scales and glottal-pulse time-scales. The top pair of rows shows a speech signal consisting of three spoken sentences (black, upper panel) decomposed into a modulator (thick positive line, upper panel) and a carrier (thin line, lower panel). The left hand panels show PAD and the right hand panels the ALP method. The time-scale used for PAD was learned from the data using a prior favouring long time-scales, and the ALP cut-off was set to give similar results to PAD. The result of demodulating the carriers (an empirical test of the Carrier Identity property) is also shown (thick line, lower panel). The bottom pair of panels shows the first syllable of the middle sentence. Here the left hand panels show PAD with a time-scale learned from the data using a prior favouring short time-scales. The right hand panels show the results from using the HE.

as is the approach to the amplitude demodulation.

V. Conclusion

This paper has introduced a new perspective on demodulation viewing it as a probabilistic inference problem. This perspective directly led to the development of an algorithm called Probabilistic Amplitude Demodulation which proceeds via an optimisation of a non-linear cost function. The Fast Fourier Transform was used to accelerate inference making PAD a real-time algorithm and further approximations, based on Laplace's method, were introduced to make learning tractable. However, despite these improvements, PAD remains computationally intensive when compared to existing approaches to demodulation. Nevertheless, PAD has several advantages. For instance, we have highlighted five desiderata which current demodulation algorithms fail to satisfy, but which are fulfilled by PAD. The first is that the method has soft constraints, which is naturally met by PAD because of the probabilistic calculus upon which it is based. Second we demonstrated that the method can automatically adjust to the signal by learning important parameters, like the time-scale of the modulator and the sparsity of the signal. Third, we have shown that the method can be steered by the user, for example on a speech signal where user-specific priors were used to select between solutions of differing time-scale. Fourth, we demonstrated that the method was robust to broadband noise added to both synthetic and natural data. And fifth, that the PAD solution was consistent, in the sense that PAD removes almost all of the modulator information from the carrier (the Carrier Identity property).

PAD not only returns an estimate of the modulator in a signal, it also returns an estimate of the uncertainty in the modulator. The fact that PAD handles uncertainties correctly, means it can be naturally extended to missing-data tasks.

One hope is that this approach can be extended to simultaneous amplitude and frequency demodulation.

Acknowledgment

The authors would like to thank the Gatsby Unit and the EPSRC for funding.

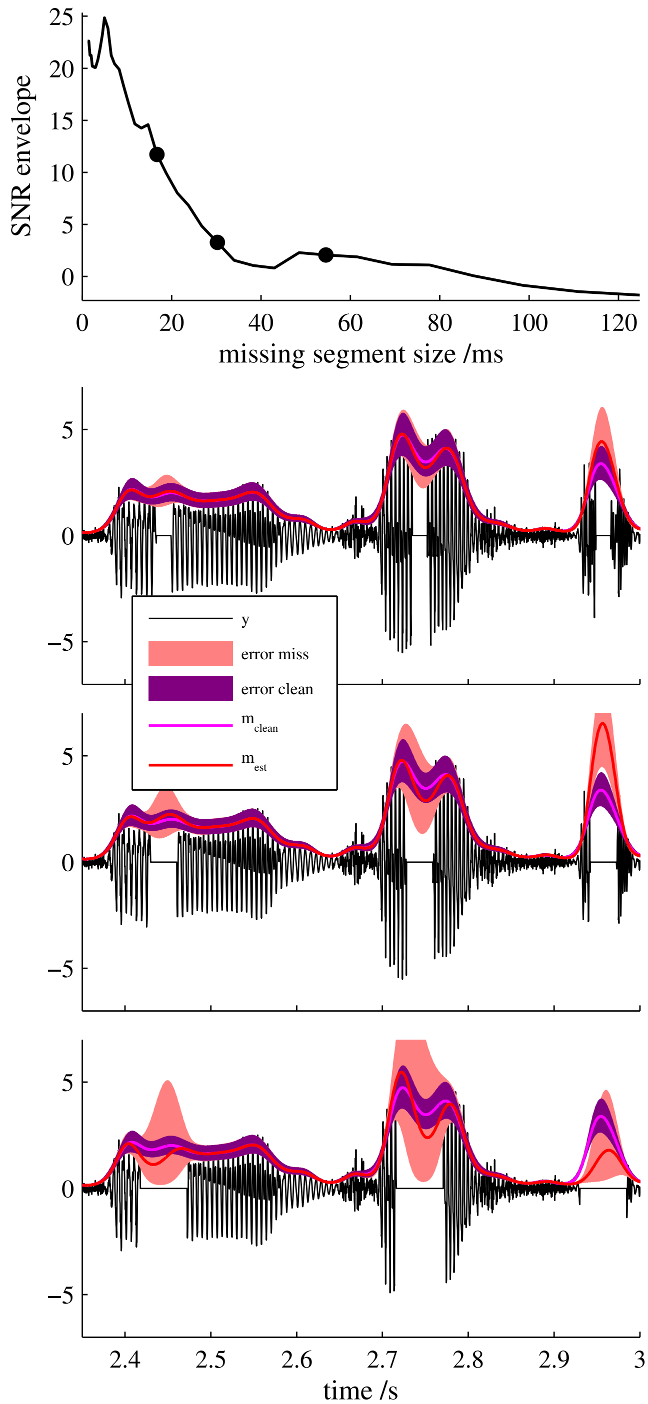

Fig. 9. Demodulating a speech signal with missing data. Short sections of the speech signal were removed from randomly chosen locations. PAD was used to fill in the missing regions and the SNR of the estimated envelopes in the missing regions was plotted as a function of gap size (top panel). The open black circles correspond to the examples plotted in the lower panels. The lower panels show short sections of the signals (black), the envelopes estimated from the clean signal with three standard deviation error-bars (dark shaded region), and the estimated envelopes with error-bars (light shaded region).

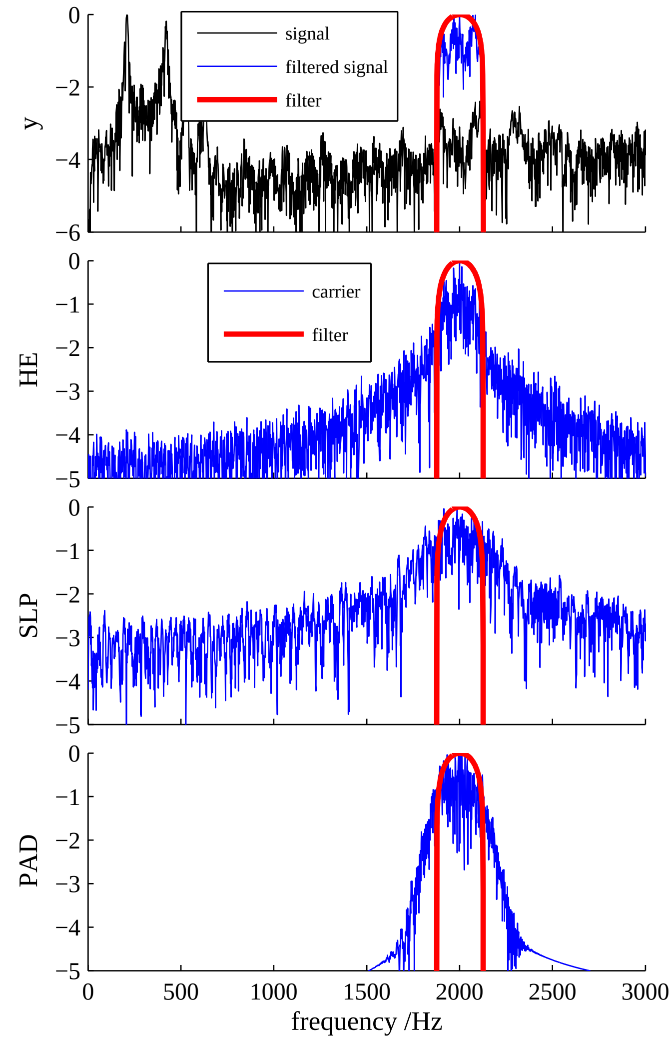

Fig. 10. Demodulating a filtered speech signal. A speech signal (top panel, normalised $\log_{10}$-spectrum shown in black) was filtered using a cosine shaped filter (centred at 2000Hz with full-width 250Hz, smooth line) and normalised to unit variance to produce a band-limited signal (top panel, within the filter pass-band). This was demodulated using the HE, SLP and PAD methods and the the lower three panels show the log-spectra of the resulting carriers. The carrier derived using PAD is closer to being limited to the pass-band of the filter than the other methods.

REFERENCES

[1] H. Dudley, "Remaking speech," *The Journal of the Acoustical Society of America*, vol. 11, no. 2, pp. 169–177, 1939.

[2] J. L. Flanagan and R. M. Golden, "Phase vocoder," *Bell System Technical Journal*, pp. 1493–1509, 1966.

[3] B. E. D. Kingsbury, N. Morgan, and S. Greenberg, "Robust speech recognition using the modulation spectrogram," *Speech Communication*, vol. 25, pp. 117–132, 1998.

[4] D. Ellis, *The Handbook of Phonetic Science*, 2nd ed. Blackwell Handbooks in Linguistics, 2008, ch. An introduction to signal processing for speech.

[5] N. Orio, *Music Retrieval: A Tutorial and Review*. now publishers Inc, 2006.

[6] S. M. Schimmel, "Theory of modulation frequency analysis and modulation filtering, with applications to hearing devices," Ph.D. dissertation, University of Washington, 2007.

[7] G. Hu, "Monaural speech segregation based on pitch tracking and amplitude modulation," *IEEE Trans. Neural Networks*, vol. 15, pp. 1135–1150, 2004. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.130.519

[8] L. Atlas and C. Janssen, "Coherent modulation spectral filtering for single-channel music source separation," in *In Proceedings of the IEEE Conference on Acoustics Speech and Signal Processing*, 2005.

[9] P. C. Loizou, M. Dorman, and Z. Tu, "On the number of channels needed

to understand speech," *Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 2097–2103, 1999.

[10] J. L. Flanagan, "Parametric coding of speech spectra," *Journal of the Acoustical Society of America*, vol. 68, pp. 412–419, 1980.

[11] R. Drullman, J. M. Festen, and R. Plomp, "Effect of temporal envelope smearing on speech reception," *The Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1053–1064, 1994.

[12] R. Shannon, F.-G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, vol. 270, pp. 303–304, 1995.

[13] M. F. Dorman, P. C. Loizou, and D. Rainey, "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," *Journal of the Acousical Society of America*, vol. 102, no. 4, pp. 2403–2411, 1997.

[14] L. Xu and B. E. Pfingst, "Relative importance of temporal envelope and fine structure in lexical-tone perception (l)," *The Journal of the Acoustical Society of America*, vol. 114, no. 6, pp. 3024–3027, 2003. [Online]. Available: http://link.aip.org/link/?JAS/114/3024/1

[15] Z. M. Smith, B. Delgutte, and A. J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," *Nature*, vol. 416, no. 6876, pp. 87–90, March 2002. [Online]. Available: http://dx.doi.org/10.1038/416087a

[16] M. G. Heinz and J. Swaminathan, "Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech." *J Assoc Res Otolaryngol*, vol. 10, no. 3, pp. 407–23, 2009. [Online]. Available: http://www.biomedsearch.com/nih/Quantifying-envelope-fine-structure-coding/19365691.html

[17] J. Dugundji, "Envelopes and pre-envelopes of real waveforms," *IEEE Transactions on Information Theory*, vol. 4, pp. 53–57, 1958.

[18] O. Ghitza, "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," *The Journal of the Acoustical Society of America*, vol. 110, no. 3, p. 16281640, 2001.

[19] F. G. Zeng, K. Nie, S. Liu, G. Stickney, E. Del Rio, Y. Y. Kong, and H. Chen, "On the dichotomy in auditory perception between temporal envelope and fine structure cues (l)," *The Journal of the Acoustical Society of America*, vol. 116, no. 3, pp. 1351–1354, 2004. [Online]. Available: http://dx.doi.org/10.1121/1.1777938

[20] P. J. Loughlin and B. Tacer, "On the amplitude- and frequency-modulation decomposition of signals," *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1594–1601, 1996. [Online]. Available: http://link.aip.org/link/?JAS/100/1594/1

[21] G. Sell and M. Slaney, "Solving demodulation as an optimization problem," *IEEE Transactions on Audio, Speech and Language Processing*, accepted.

[22] P. Clark and L. E. Atlas, "Time-frequency coherent modulation filtering of nonstationary signals," *Trans. Sig. Proc.*, vol. 57, no. 11, pp. 4323–4332, 2009.

[23] R. Libbey, *Signal and image processing sourcebook*. Springer, 1994.

[24] D. Vakman, "On the analytic signal, the teager-kaiser energy algorithm, and other methods for defining amplitude and frequency," *IEEE Journal of Signal Processing*, vol. 44, no. 4, pp. 791–797, 1996.

[25] L. Atlas, Q. Li, and J. Thompson, "Homomorphic modulation spectra," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, 2004, pp. 17–21.

[26] R. E. Turner and M. Sahani, "Probabilistic amplitude demodulation," in *Independent Component Analysis and Signal Separation*, 2007, pp. 544–551, best student paper award.

[27] *Statistical inference for single- and multi-band probabilistic amplitude demodulation.*, 2010.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, March 2004. [Online]. Available: http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0521833787

[29] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003, available from http://www.inference.phy.cam.ac.uk/mackay/itila/. [Online]. Available: citeseer.ist.psu.edu/mackay03information.html

[30] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, December 2006.

[31] R. E. Turner, "Statistical models for natural sounds," Ph.D. dissertation, Gatsby Computational Neuroscience Unit, UCL, 2009.

[32] K. Atkinson, *An Introduction to Numerical Analysis*. John Wiley and Sons, 1988.

[33] A. Bultheel and M. V. Barel, "Chapter 3 lanczos algorithm," in *Linear Algebra, Rational Approximation and Orthogonal Polynomials*, ser. Studies in Computational Mathematics. Elsevier, 1997, vol. 6, pp. 99 – 133. [Online]. Available: http://www.sciencedirect.com/science/article/B8GXW-4NXP6D7-5/2/2bed37eb1b7f32aad12c63cf0beccf32