VARIATIONAL LOG-POWER SPECTRAL TRACKING FOR ACOUSTIC SIGNALS

Bart van Erp[⋆] *İsmail Şenöz*[⋆] *Bert de Vries*[⋆†]

[⋆] Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven (NL)
[†] GN Hearing, Eindhoven (NL)

ABSTRACT

This paper proposes a generative hierarchical probabilistic model for acoustic signals where both the frequency decomposition and log-power spectrum appear as latent variables. In order to facilitate efficient inference, we represent the model in a factor graph that includes a probabilistic Fourier transform and a Gaussian scale model as modules. We derive novel ways of performing variational message passing-based inference in the Gaussian scale model. As a result, in this model a probabilistic representation of the log-power spectrum of an acoustic signal can be effectively inferred online. The proposed model may find applications as a front end wherever probabilistic log-power spectral features of a signal are needed. We validate the model and message passing-based inference methods by tracking the log-power spectrum of a speech signal.

1. INTRODUCTION

Various probabilistic generative models for different representations of acoustic signals have been presented in the literature. For example, [1] models the temporal representation using a probabilistic auto-regressive model, [2, 3] model the frequency coefficients independently using probabilistic auto-regressive models and [4, 5, 6, 7] model the log-power spectrum using Gaussian mixture models. Despite these independent modeling strategies, a principled joint spectro-temporal modeling approach has not yet been proposed.

During the probabilistic modeling of acoustic signals, the frequency coefficients or log-power spectra are often first extracted from the signal. A less common approach is to treat these different representations as latent states in a probabilistic model. In [8] the probabilistic version of the Fourier transform is introduced, stochastically relating the observed time domain signal to its frequency coefficients, which now represent hidden random variables. In [9] and [2] two models using this approach are presented. These are closely related to each other [3] and to Gaussian processes [10].

Similarly, in [11] the relation between the log-power spectrum and frequency coefficients is modeled as a probability density function, termed the Gaussian scale model. Here the frequency coefficients are modelled by a complex Gaussian distribution, whose covariances are exponentiated random variables, representing the log-power spectrum. [11] specifically models this probabilistic log-power spectrum by a Gaussian mixture model, resulting in a Gaussian scale mixture model (GSMM) [12]. However, a general approach for performing inference in the Gaussian scale model is missing, restricting implementation with alternative "back ends" for modeling the probabilistic log-power spectrum.

This paper combines the probabilistic Fourier transform and Gaussian scale model in a hierarchical probabilistic "front end" for more complex probabilistic models. The front end we propose jointly models the time domain, frequency coefficients and log-power spectrum of an acoustic signal, based on a noisy temporal observation. Importantly, in our approach the different representations are treated as latent variables in the generative model. Furthermore, we propose several ways for performing probabilistic inference in this hierarchical front end using message passing.

This paper is organized as follows. In Section 2 we propose a probabilistic generative model for acoustic signals. In order to support inference in this model, we will represent the model by a Forney-style factor graph (FFG). Message passing-based inference in FFGs is shortly reviewed in Section 3. Section 4 describes how probabilistic inference for log-power spectral tracking in our model can be performed using message passing. We evaluate our approach in Section 5 and conclude the paper in Section 6.

Correspondence: b.v.erp@tue.nl

2. MODEL SPECIFICATION

This section specifies our probabilistic generative model for acoustic signals. Consider an observed signal $\boldsymbol{y}_n = [y_n, y_{n-1}, \dots y_{n-N}]^\top$ of length $N+1$ at time $t = nT$, where T is the sampling period and n the sample index. We assume that $\boldsymbol{y}_n$ is composed of a "clean" signal $\boldsymbol{x}_n = [x_n, x_{n-1}, \dots x_{n-N}]^\top$ plus independent and identically distributed (i.i.d.) Gaussian observation noise. The clean time domain signal $\boldsymbol{x}_n$ relates to its frequency coefficients through

$$\boldsymbol{x}_n = F_n \boldsymbol{s}_n, \tag{1}$$

where F_n is defined as

$$F_n = \begin{bmatrix} \cos(\omega_1 nT) & \cdots & \cos(\omega_1 (n-N)T) \\ \vdots & & \vdots \\ \cos(\omega_M nT) & \cdots & \cos(\omega_M (n-N)T) \\ \sin(\omega_1 nT) & \cdots & \sin(\omega_1 (n-N)T) \\ \vdots & & \vdots \\ \sin(\omega_M nT) & \cdots & \sin(\omega_M (n-N)T) \end{bmatrix}^\top \tag{2}$$

and where $\boldsymbol{s}_n = [s_n^1, s_n^2, \dots, s_n^{2M}]^\top$ represents a vector of frequency coefficients s_n^m with angular frequency ω_m at time index n. Importantly, both $\boldsymbol{x}_n$ and $\boldsymbol{s}_n$ are random vectors in (1). The probabilistic Fourier transform [8] is here expressed as the likelihood function

$$p(\boldsymbol{y}_n \mid \boldsymbol{s}_n) = \mathcal{N}\left(\boldsymbol{y}_n \mid F_n \boldsymbol{s}_n,\ \Lambda_y^{-1}\right), \tag{3}$$

where Λ_y is a diagonal precision matrix for the observation noise.

The resulting vector of real coefficients $\boldsymbol{s}_n$ can be reparameterized into a vector of complex frequency coefficients $\boldsymbol{c}_n$ through

$$c_n^m = s_n^m + i \cdot s_n^{m+M} \quad \text{for } m = 1, \dots, M \tag{4}$$

where $i = \sqrt{-1}$. This reordering is just a reflection of the fact that the first half of the entries of $\boldsymbol{s}_n$ corresponds to the real parts of the corresponding complex frequency coefficients and the second half corresponds to the imaginary parts of these coefficients.

We will assume a (complex) Gaussian distribution for $\boldsymbol{c}_n$, given by $\mathcal{N}_{\mathbb{C}}(\boldsymbol{c}_n \mid \boldsymbol{\mu}, \Gamma, C)$ with mean $\boldsymbol{\mu} = 0$, complex covariance matrix Γ and relation matrix C, see [13] for more details. In order to keep inference tractable, independence is assumed between the real and imaginary parts of the coefficients, requiring $C = 0$. Following [11], the covariance matrix Γ is modelled as a diagonal matrix with exponentiated auxiliary variables ξ_n^m, leading to

$$p(c_n^m \mid \xi_n^m) = \mathcal{N}_{\mathbb{C}}\left(c_n^m \mid 0,\ e^{\xi_n^m},\ 0\right) = \frac{e^{-\xi_n^m}}{\pi} e^{-e^{-\xi_n^m}|c_n^m|^2}. \quad (5)$$

This probabilistic relationship relates the complex frequency coefficients to the log-power spectrum [11]. To clarify this point, the log-likelihood function of (5) can be found as $\ln p(c_n^m \mid \xi_n^m) = -\xi_n^m - e^{-\xi_n^m}|c_n^m|^2 - \ln \pi$ and from this description the maximum of ξ_n^m can be found to occur at $\xi_n^m = \ln |c_n^m|^2$, which coincides with the deterministic transform from the frequency coefficients to the log-power spectrum. As a result of this observation, the vector $\boldsymbol{\xi}_n = [\xi_n^1, \xi_n^2, \ldots, \xi_n^M]^\top$ is treated as the probabilistic log-power spectrum of the noisy acoustic signal $\boldsymbol{y}_n$.

We model the log-power spectrum as a Gaussian random walk

$$p\left(\xi_n^m \mid \xi_{n-K}^m\right) = \mathcal{N}\left(\xi_n^m \mid \xi_{n-K}^m,\ \gamma_\xi^{-1}\right), \quad (6)$$

where we model frames of length $N+1$ with a step size K between frames and where γ_ξ represents the process noise precision. Because of the specified step size, ξ_n^m does not have to be defined for all n. In principle, it would be easy to replace the Gaussian random walk with an alternative model such as a probabilistic auto-regressive model, for which message passing update rules have been derived in [1].

The set of equations (3), (4), (5) and (6) specifies a generative probabilistic model for acoustic signals that includes both the frequency decomposition and log-power spectrum as latent variables. In general, we observe signals in the temporal domain and we are interested in inferring the log-power spectrum representation (or the frequency decomposition). Inference in this model is not analytically tractable. We solve the inference task through message passing in a factor graph and for this purpose we derive new variational message passing update rules. In the next section, we shortly summarize message passing-based inference in factor graphs.

3. FACTOR GRAPHS AND MESSAGE PASSING

We use message passing in a factor graph as our probabilistic inference approach of choice, because of its efficiency, automatability, scalability and modularity [14, 15].

3.1. Forney-style factor graphs

Factor graphs are a class of probabilistic graphical models. This paper will discuss Forney-style factor graphs (FFG) as introduced in [16] with notational conventions adopted from [17]. The interested reader may refer to [17] or [14] for additional information on FFGs. FFGs visualize global factorizable functions as an undirected graph of nodes corresponding to the local functions, or factors, connected by edges representing their mutual arguments. This factorized representation allows naturally for the visualization of conditional dependencies in generative probabilistic models. Figure 1 shows a factor graph representation of a single time slice of our generative model.

3.2. Sum-product message passing

For calculating marginal distributions in a generative model, we need to integrate over all other random variables. Because of the factorization of the probabilistic model, we can perform this marginalization through a set of small local computations. These local computations are called messages and are denoted by μ. The sum-product message $\vec{\mu}(x_j)$ flowing out of an arbitrary node $f(x_1, x_2, \ldots, x_n)$ with incoming messages $\vec{\mu}(x_{\setminus j})$ is given by

$$\vec{\mu}(x_j) = \int f(x_1, x_2, \ldots, x_n) \prod_{i \neq j} \vec{\mu}(x_i)\, \mathrm{d}\boldsymbol{x}_{\setminus j} \quad (7)$$

which is called the sum-product update rule [18]. This update rule is the core of the sum-product message passing algorithm, which is also known as belief propagation [19]. This algorithm concerns the distributed calculation of various marginal functions from a factorizable global function. The marginal distributions can then be calculated from the messages as $p(x_j) \propto \vec{\mu}(x_j) \cdot \overleftarrow{\mu}(x_j)$. The FFG now has arbitrarily directed edges to indicate the flow of the messages. A message $\mu(x_j)$ propagating on edge x is denoted by $\vec{\mu}(x_j)$ or $\overleftarrow{\mu}(x_j)$ when propagating in or against the direction of the edge, respectively.

3.3. Variational message passing

In some cases the integrals in the sum-product algorithm can become intractable. Then we can resort to an approximate message passing algorithm, called variational message passing (VMP) [20], [21].

Consider the generative model $p(y, x)$ with an intractable posterior distribution $p(x|y)$, where y and x denote the observed and latent variables, respectively. Variational inference approximates the intractable true posterior with a tractable variational distribution $q(x)$ through minimization of a variational free energy functional

$$F[q] = \underbrace{\int q(x) \ln \frac{q(x)}{p(x|y)}\, \mathrm{d}x}_{\text{KL-divergence}} - \underbrace{\ln p(y)}_{\text{log-evidence}} \quad (8)$$

which is in the machine learning literature also known as the negative Evidence Lower BOund (ELBO).

In practice the optimization of (8) is performed by imposing additional constraints on $q(x)$, e.g., by limiting $q(x)$ to a family of distributions, or by additional factorization assumptions. Depending on the constraints on $q(x)$, the minimization of (8) can be achieved through sum-product message passing or variants of VMP. In the latter case, the goal is to iteratively update the variational distributions through coordinate descent on (8). In general, the variational message $\nu(x_j)$ from a generic node $f(x_1, x_2, \ldots, x_n)$ with incoming messages $q(x_{\setminus j})$ can be written as [21]

$$\vec{\nu}(x_j) \propto \exp \int \prod_{i \neq j} q(x_i) \ln f(x_1, x_2, \ldots, x_n)\, \mathrm{d}\boldsymbol{x}_{\setminus j}. \quad (9)$$

The posterior $q(x_j)$ can be updated by multiplying the forward and backward messages on the edge of x_j as $q(x_j) \propto \vec{\nu}(x_j) \cdot \overleftarrow{\nu}(x_j)$.

4. VARIATIONAL LOG-POWER SPECTRAL TRACKING

This section describes the inference procedure in the probabilistic front end using message passing. Specifically, this section describes the inference procedure in the Gaussian scale model of (5), as all other messages have already been derived in [17] and [14]. In order to prevent notational clutter, we drop the sub- and superscripts in (5).

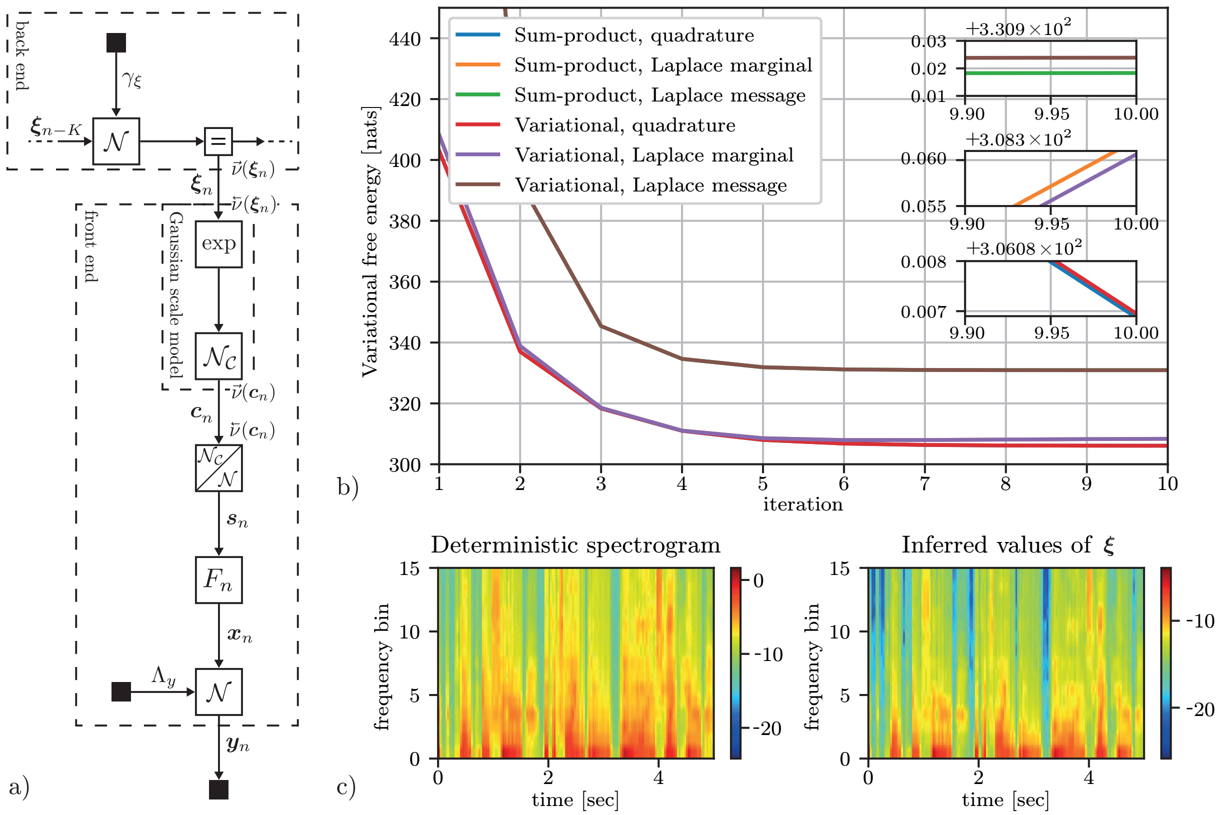

Fig. 1. This figure shows: a) a factor graph representation of a time slice of the generative model of Section 2, b) a comparison between the different inference approaches expressed as the variational free energy as a function of coordinate descent iterations, c) the deterministic and inferred log-power spectrum of an acoustic signal.

4.1. Sum-product messages

The sum-product messages can be calculated using (7), where we assume that the incoming messages to the node of (5) are given by

$$\vec{\mu}(\xi) \propto \mathcal{N}(\xi \mid m_\xi,\, \gamma_\xi^{-1}) \tag{10a}$$

$$\vec{\mu}(c) \propto \mathcal{N}_{\mathcal{C}}(c \mid m_c,\, \gamma_c^{-1},\, 0)\,, \tag{10b}$$

with means $m_\xi \in \mathbb{R}$ and $m_c \in \mathbb{C}$ and real precisions $\{\gamma_\xi, \gamma_c\} \in \mathbb{R}^+$. Substitution of (5) and (10b) in (7) leads to the message

$$\overleftarrow{\mu}(\xi) \propto \frac{e^{-\xi}}{e^{-\xi}+\gamma_c} \cdot \exp\left\{\frac{\gamma_c^2 |m_c|^2}{e^{-\xi}+\gamma_c}\right\}. \tag{11}$$

This message does not belong to the exponential family and multiplication of $\overleftarrow{\mu}(\xi)$ with $\vec{\mu}(\xi)$ will no longer be a conjugate operation. Section 4.3 will describe ways of dealing with these messages. Unfortunately, the integral involving the calculation of $\vec{\mu}(c)$ is intractable. Because of this, we need to resort to VMP [21].

4.2. Variational messages

For mean-field variational message passing we will assume the following form constraints on the posterior distributions

$$q(\xi) = \mathcal{N}(\xi \mid m_\xi,\, \gamma_\xi^{-1}) \tag{12a}$$

$$q(c) = \mathcal{N}_{\mathcal{C}}(c \mid m_c,\, \gamma_c^{-1},\, 0)\,, \tag{12b}$$

where $m_\xi \in \mathbb{R}$ and $m_c \in \mathbb{C}$ are means and $\{\gamma_\xi, \gamma_c\} \in \mathbb{R}^+$ are precisions. Substitution of (5) and (12a) in variational update rule (9) yields

$$\vec{\nu}(c) \propto \mathcal{N}_{\mathcal{C}}(0,\, \exp(m_\xi - \gamma_\xi^{-1}/2),\, 0) \tag{13}$$

and similarly substitution of (5) and (12b) in (9) yields

$$\overleftarrow{\nu}(\xi) \propto \exp\left\{-\xi - e^{-\xi}(\gamma_c^{-1} + |m_c|^2)\right\}. \tag{14}$$

This latter message again has a functional form, similar to (11).

4.3. Handling non-conjugate messages with a functional form

The messages derived in (11) and (14) do not belong to a known family of distributions and are therefore represented in their functional form. Furthermore, multiplications with these messages will no longer be conjugate operations. Here we will describe 3 ways of dealing with these messages based on [22] and [23].

First, we can approximate these messages by a Gaussian distribution directly using Laplace's method. Here the log-message is approximated by a second-order Taylor expansion at its mode as

$$\ln \overleftarrow{\nu}(\xi) \approx \ln \overleftarrow{\nu}(\xi_0) + \frac{1}{2}\frac{\mathrm{d}^2 \ln \overleftarrow{\nu}(\xi)}{\mathrm{d}\xi^2}\Big|_{\xi=\xi_0} (\xi-\xi_0)^2 \tag{15}$$

where ξ_0 is the mode of the message. Because the message is expanded around its mode, the first-order derivative vanishes from the Taylor expansion. This mode can be found by solving $\frac{\mathrm{d}\ln \overleftarrow{\nu}(\xi)}{\mathrm{d}\xi} = 0$

for ξ. This approach results in the approximate variational messages

$$\overleftarrow{\mu}(\xi) \propto \mathcal{N}(\hat{m}_\xi,\ \hat{\gamma}_\xi^{-1}) \tag{16a}$$

$$\overleftarrow{\nu}(\xi) \propto \mathcal{N}(\ln(\gamma_c^{-1} + |m_c|^2),\ 1), \tag{16b}$$

where

$$\hat{m}_\xi = -\ln\left(\frac{\gamma_c}{\gamma_c|m_c|^2 - 1}\right)$$

$$\hat{\gamma}_\xi = \frac{-e^{-\hat{m}_\xi}}{\left(e^{-\hat{m}_\xi} + \gamma_c\right)^2}\left(-\gamma_c - \gamma_c^2|m_c|^2 + \frac{2\gamma_c^2|m_c|^2 e^{-\hat{m}_\xi}}{e^{-\hat{m}_\xi} + \gamma_c}\right).$$

A second approach is to propagate the message on the graph and to approximate the resulting marginal instead of the message itself. Here the marginal distribution $q(\xi)$ is given by $q(\xi) \propto \vec{\nu}(\xi) \cdot \overleftarrow{\nu}(\xi)$, where one of the messages is a function and the other a Gaussian. The Gaussian can also be converted to its functional form and the marginal can be expressed as the product of the colliding messages. The mode of the resulting marginal can be approximated using Newton's method, where the first and second derivatives can be determined using automatic differentiation. Similarly the variance of the approximate Gaussian marginal distribution can be determined.

The third approach uses moment matching instead of Laplace's method for approximating the resulting marginals. The moments of the marginal $q(\xi)$ can be calculated as

$$\mathbb{E}_{q(\xi)}[\xi^n] = \int \xi^n \frac{\vec{\nu}(\xi)\overleftarrow{\nu}(\xi)}{\int \vec{\nu}(\xi)\overleftarrow{\nu}(\xi)\mathrm{d}\xi}\mathrm{d}\xi \tag{17}$$

in which the marginal is properly normalized by dividing by the normalization constant $Z = \int \vec{\nu}(\xi)\overleftarrow{\nu}(\xi)\mathrm{d}\xi$. This intractable normalization integral is can be approximated using Gauss-Hermite quadrature integration [24, Chapter 6] as

$$Z \approx \frac{1}{\sqrt{\pi}}\sum_{k=1}^{K} w_k \overleftarrow{\nu}\left(\bar{\xi}_k\sqrt{2\gamma_\xi^{-1}} + m_\xi\right), \tag{18}$$

where w_k and $\bar{\xi}_k$ can be calculated using the Golub-Welsch algorithm [25], representing the weights and evaluation points, respectively. Similarly, the moments can be approximated using Gauss-Hermite quadrature integration as

$$\mathbb{E}[\xi^n] \approx \frac{1}{Z\sqrt{\pi}}\sum_{k=1}^{K} w_k\left(\bar{\xi}_k\sqrt{2\gamma_\xi^{-1}} + m_\xi\right)^k \overleftarrow{\nu}\left(\bar{\xi}_k\sqrt{2\gamma_\xi^{-1}} + m_\xi\right)$$

and from the obtained moments, the sufficient statistics of the approximate Gaussian marginal distribution can be found.

5. EXPERIMENTAL VALIDATION

5.1. Log-power spectral tracking of acoustic signals

In order to validate the proposed methodology, the model is used to perform generalized Kalman smoothing to estimate the probabilistic log-power spectrum of an acoustic signal.[1] A random fragment of female speech lasting 5 seconds and resampled to $f_s = 16$ kHz has been selected. We process the time domain signal in blocks of length $N + 1 = 32$ with a step size of $K = 32$. From these blocks the average value is subtracted. For the frequency decomposition the angular frequencies $\omega_m = m\frac{2\pi f_s}{N+1}$ for $m = 1, \ldots, (N-1)/2$ are used. In the proposed model we set the hyperparameters as $\Lambda_y = 10^8 \cdot \mathrm{I}_{N+1}$ and $\gamma_\xi = 10^{-3}$, where I_{N+1} denotes the identity matrix of size $(N+1 \times N+1)$. The choice of these parameters illustrates a model in which we assume little observation noise and a log-power spectrum with little temporal dependency.

Technically, the log-power spectral estimation task can be phrased as computing $q(\boldsymbol{\xi}_{1:K:n}|\boldsymbol{y}_{1:K:n})$. For validation of our approach, we will compare the obtained results with the deterministic log-power spectrum. The open source Julia [26] package ForneyLab is used, which aims to excel at real-time message passing-based inference in dynamic models [15]. The inference in the Gaussian scale model is performed using VMP, where the message $\overleftarrow{\nu}(\xi)$ is approximated directly using Laplace's method by (16b). Fig. 1 shows a comparison between the deterministic log-power spectrum and the smoothed probabilistic log-power spectrum.

5.2. Comparison between inference methods

Fig. 1 also shows a comparison between the different inference approaches proposed for the Gaussian scale model. For this purpose we directly model the complex frequency coefficients using a simplified generative model, consisting out of observation model $p(z_n \mid c_n) = \mathcal{N}(z_n \mid c_n, \gamma_c^{-1})$, (5) and (6) for a single frequency coefficient and unit time steps, meaning that $M = 1$ and $K = 1$. Here z_n represents the observed complex frequency coefficients and γ_c represents the observation noise precision. Synthetic data is used for comparison, consisting out of 100 data points, generated using a process noise precision of $\gamma_\xi = 1$ and observation noise precision of $\gamma_c = 10^6$. The prior $p(\xi_0)$ equals the standard Gaussian distribution. Fig. 1 compares the performance of the different inference methods through the variational free energy.

5.3. Discussion

From Fig. 1 we can conclude that the proposed front-end is capable of extracting the spectral characteristics of an acoustic signal, solely through probabilistic inference. The process noise precision controls the smoothing of $\boldsymbol{\xi}$ and is kept low for comparison sake. The approaches where the marginal distributions are approximated using Laplace's method or Gauss-Hermite quadrature integration outperform the approaches where the message is approximated using Laplace's method in terms of variational free energy.

The current proposed approach allows us to model all representations of an acoustic signal simultaneously. Furthermore it allows us to explicitly measure information loss in transforming the acoustic signal by the resulting variational free energy. Finally, the current front-end provides a modular basis for alternative model assumptions on the probabilistic log-power spectrum.

6. CONCLUSION

This paper presented a generative probabilistic model for acoustic signals where the frequency decomposition and log-power spectrum are represented as latent model variables. In order to facilitate online inference, we represent our model as a factor graph and derive different variational update rules for the Gaussian scale model, which is a sub-module in our model. Experimental validation supports the notion that we proposed an efficient variational log-power spectral tracking algorithm that can be applied as a plug-in front end to various probabilistic signal processing tasks.

[1]Experiments are available at `https://github.com/biaslab/SSP2021-VariationalLogPowerTracking`.

7. REFERENCES

[1] A. Podusenko, W. M. Kouw, and B. de Vries, "Online Variational Message Passing in Hierarchical Autoregressive Models," in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 1337–1342, June 2020. ISSN: 2157-8117.

[2] A. T. Cemgil and S. J. Godsill, "Probabilistic phase vocoder and its application to interpolation of missing values in audio signals," in *2005 13th European Signal Processing Conference*, pp. 1–4, Sept. 2005.

[3] R. E. Turner and M. Sahani, "Time-Frequency Analysis as Probabilistic Inference," *IEEE Transactions on Signal Processing*, vol. 62, pp. 6171–6183, Dec. 2014.

[4] S. Rennie, J. Hershey, and P. Olsen, "Single-channel speech separation and recognition using loopy belief propagation," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009*, pp. 3845–3848, Apr. 2009.

[5] S. Rennie, J. Hershey, and P. Olsen, "Single-Channel Multitalker Speech Recognition," *IEEE Signal Processing Magazine*, vol. 27, pp. 66–80, Nov. 2010.

[6] B. J. Frey, L. Deng, A. Acero, and T. Kristjansson, "ALGONQUIN: Iterating Laplace's Method to Remove Multiple Types of Acoustic Distortion for Robust Speech Recognition," *Proc. of the Eurospeech Conference*, pp. 901–904, Sept. 2001.

[7] J. R. Hershey, P. Olsen, and S. J. Rennie, "Signal Interaction and the Devil Function," *INTERSPEECH-2010*, pp. 334–337, 2010.

[8] G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*. Lecture Notes in Statistics, New York, NY: Springer, 1988.

[9] Y. Qi, T. P. Minka, and R. W. Picara, "Bayesian spectrum estimation of unevenly sampled nonstationary data," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. II–1473–II–1476, May 2002. ISSN: 1520-6149.

[10] W. J. Wilkinson, M. R. Andersen, J. D. Reiss, D. Stowell, and A. Solin, "Unifying Probabilistic Models for Time-Frequency Analysis," *arXiv:1811.02489 [cs, eess, stat]*, Feb. 2019.

[11] J. Hao, T.-W. Lee, and T. J. Sejnowski, "Speech Enhancement Using Gaussian Scale Mixture Models," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, pp. 1127–1136, Aug. 2010.

[12] D. F. Andrews and C. L. Mallows, "Scale Mixtures of Normal Distributions," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 36, no. 1, pp. 99–102, 1974. Publisher: [Royal Statistical Society, Wiley].

[13] R. G. Gallager, "Circularly-Symmetric Gaussian random vectors," *preprint*, Jan. 2008. Accessed from: https://www.rle.mit.edu/rgallager/documents/CircSymGauss.pdf.

[14] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The Factor Graph Approach to Model-Based Signal Processing," *Proceedings of the IEEE*, vol. 95, pp. 1295–1322, June 2007.

[15] M. Cox, T. van de Laar, and B. de Vries, "A Factor Graph Approach to Automated Design of Bayesian Signal Processing Algorithms," *International Journal of Approximate Reasoning*, vol. 104, pp. 185–204, Jan. 2019. arXiv: 1811.03407.

[16] G. Forney, "Codes on graphs: normal realizations," *IEEE Transactions on Information Theory*, vol. 47, pp. 520–548, Feb. 2001.

[17] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Magazine*, vol. 21, pp. 28–41, Jan. 2004.

[18] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, pp. 498–519, Feb. 2001.

[19] J. Pearl, "Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach," in *Proceedings of the American Association for Artificial Intelligence National Conference on AI*, (Pittsburgh), pp. 133–136, 1982.

[20] J. Winn and C. M. Bishop, "Variational Message Passing," *Journal of Machine Learning Research*, pp. 661–694, 2005.

[21] J. Dauwels, "On Variational Message Passing on Factor Graphs," in *2007 IEEE International Symposium on Information Theory*, pp. 2546–2550, June 2007. ISSN: 2157-8117.

[22] İ. Şenöz and B. de Vries, "Online Variational Message Passing in the Hierarchical Gaussian Filter," in *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, Sept. 2018. ISSN: 1551-2541.

[23] İ. Şenöz and B. de Vries, "Online Message Passing-based Inference in the Hierarchical Gaussian Filter," in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 2676–2681, June 2020. ISSN: 2157-8117.

[24] S. Särkkä, *Bayesian Filtering and Smoothing*. Institute of Mathematical Statistics Textbooks, Cambridge: Cambridge University Press, 2013.

[25] G. H. Golub and J. H. Welsch, "Calculation of Gauss quadrature rules," *Mathematics of Computation*, vol. 23, no. 106, pp. 221–230, 1969.

[26] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A Fresh Approach to Numerical Computing," *SIAM Review*, vol. 59, pp. 65–98, Jan. 2017. Publisher: Society for Industrial and Applied Mathematics.