Supplementary Material for
End-to-End Probabilistic Inference for Nonstationary Audio Analysis

In this appendix we show further details regarding the globally iterated extended Kalman filter algorithm and the derivations for moment matching in expectation propagation.

A. Global Iterated Extended Kalman Filter

Alg. 2 provides pseudo code for the inference algorithm that uses the classical extended Kalman filtering method.

Algorithm 2 Linearisation-based inference (Laplace approximation scheme) formulated as a global iterated extended Kalman filter.

```
Input: {t_k}, y, A, Q, P_0                          data and state space model
       h(x), H_x(x)                                 measurement model and Jacobian
m_0 ← 0                                             init state mean
while not converged do                              iterated EKF loop
  for k = 1 to T do                                 forward pass
    if k == 1 then
      P_k ← P_0                                     init state covariance
    else
      m_k ← A m_{k-1}; P_k ← A P_{k-1} A^⊤ + Q      predict
    end if
    if has label y_k then
      v_k ← y_k − h(m_k); S_k ← H_x P_k H_x^⊤ + σ_y^2     inn.
      k_k ← P_k H_x^⊤ S_k^{-1}                      gain
      m_k ← m_k + k_k v_k; P_k ← P_k − k_k S_k k_k^⊤
    end if
  end for
  for k = T − 1 to 1 do                             backward pass
    G_k ← P_k A^⊤ (A P_k A^⊤ + Q)^{-1}              gain
    m_k ← m_k + G_k (m_{k+1} − A m_k)
    P_k ← P_k + G_k (P_{k+1} − A P_k A^⊤ − Q) G_k^⊤
  end for
end while
  measurement row vector h_n^g selects g_n from the state
Return: E[g_n(t_k)] = h_n^g m_k; V[g_n(t_k)] = h_n^g P_k h_n^g⊤
        E[z_d(t_k)] = h_d^z m_k; V[z_d(t_k)] = h_d^z P_k h_d^z⊤
        log p(y | θ) ≃ − Σ_{k=1}^T 1/2 (log 2π S_k + v_k^2 / S_k)
```

B. Additional Derivations for EP

The normalisation constant required for calculating the posterior in GTF-NMF, as described in Sec. 3.4, is:

$$
\begin{aligned}
Z_k = \;& \mathbb{E}_{q_{-k}}\left[p(y_k \,|\, \mathbf{g}_k, \mathbf{z}_k)^\eta\right] \\
= \;& \int ... \int \mathrm{N}\Big(y_k \,|\, \sum_d \sum_n W_{d,n}\psi(g_{n,k}) z_{d,k}, \sigma_y^2\Big)^\eta \\
& \times \prod_d \mathrm{N}\big(z_{d,k} \,|\, \mu^z_{d,-k}, \zeta^z_{d,-k}\big) \\
& \times \prod_n \mathrm{N}\big(g_{n,k} \,|\, \mu^g_{n,-k}, \zeta^g_{n,-k}\big) \mathrm{d}z_{1,k}...\mathrm{d}z_{D,k}\mathrm{d}g_{1,k}...\mathrm{d}g_{N,k} \\
= \;& \mathrm{const}_\eta \int ... \int \mathrm{N}\Big(y_k \,|\, \sum_d \sum_n W_{d,n}\psi(g_{n,k})\mu^z_{d,-k}, \\
& \sigma_y^2 + \sum_d \sum_n W^2_{d,n}\psi(g_{n,k})^2 \zeta^z_{d,-k}\Big) \\
& \times \prod_n \mathrm{N}\big(g_{n,k} \,|\, \mu^g_{n,-k}, \zeta^g_{n,-k}\big) \mathrm{d}g_{1,k}...\mathrm{d}g_{N,k}
\end{aligned}
$$

for $\mathrm{const}_\eta = (2\pi\sigma_y^2)^{1/2(1-\eta)}\eta^{-1/2}$ and where we have used the marginalisation properties of the Gaussian distribution to obtain the last line.

Setting

$$
m_y = \sum_d \sum_n W_{d,n}\psi(g_{n,k})\mu^z_{d,-k}
$$

and

$$
v_y = \sigma_y^2 + \sum_d \sum_n W^2_{d,n}\psi(g_{n,k})^2 \zeta^z_{d,-k}
$$

and differentiating w.r.t. μ^z, μ^g, we get

$$
\begin{aligned}
\frac{\mathrm{d}Z_k}{\mathrm{d}\mu^z_{d,-k}} = \;& \mathrm{const}_\eta \int ... \int \mathrm{N}\big(y_k \,|\, m_y, v_y\big) \\
& \times \sum_n W_{d,n}\psi(g_{n,k}) \frac{y - m_y}{v_y} \\
& \times \prod_n \mathrm{N}\big(g_{n,k} \,|\, \mu^g_{n,-k}, \zeta^g_{n,-k}\big) \mathrm{d}g_{1,k}...\mathrm{d}g_{N,k},
\end{aligned}
$$

$$\frac{\mathrm{d}Z_k}{\mathrm{d}\mu^g_{n,-k}} = \text{const}_\eta \int ... \int \mathrm{N}\big(y_k \,|\, m_y, v_y\big) \frac{g_{n,k} - \mu^g_{n,-k}}{\zeta^g_{n,-k}} \times \prod_n \mathrm{N}\big(g_{n,k} \,|\, \mu^g_{n,-k}, \zeta^g_{n,-k}\big) \mathrm{d}g_{1,k} ... \mathrm{d}g_{N,k},$$

$$\frac{\mathrm{d}^2 Z_k}{\mathrm{d}\mu^z_{d,-k}{}^2} = \text{const}_\eta \int ... \int \mathrm{N}\big(y_k \,|\, m_y, v_y\big) \times \sum_n \left(W_{d,n}\psi(g_{n,k})\right)^2 \left[\left(\frac{y - m_y}{v_y}\right)^2 - \frac{1}{v_y}\right] \times \prod_n \mathrm{N}\big(g_{n,k} \,|\, \mu^g_{n,-k}, \zeta^g_{n,-k}\big) \mathrm{d}g_{1,k} ... \mathrm{d}g_{N,k},$$

$$\frac{\mathrm{d}^2 Z_k}{\mathrm{d}\mu^g_{n,-k}{}^2} = \text{const}_\eta \int ... \int \mathrm{N}\big(y_k \,|\, m_y, v_y\big) \times \left[\left(\frac{g_{n,k} - \mu^g_{n,-k}}{\zeta^g_{n,-k}}\right)^2 - \frac{1}{\zeta^g_{n,-k}}\right] \times \prod_n \mathrm{N}\big(g_{n,k} \,|\, \mu^g_{n,-k}, \zeta^g_{n,-k}\big) \mathrm{d}g_{1,k} ... \mathrm{d}g_{N,k}.$$

We can see from the above that all the required integrals are N-dimensional, where N is the number of NMF components.

The partial derivatives of $\log Z_k$ can be obtained from the equations above using the chain rule:

$$\frac{\mathrm{dlog}\, Z_k}{\mathrm{d}\mu^z_{d,-k}} = \frac{1}{Z_k} \frac{\mathrm{d}Z_k}{\mathrm{d}\mu^z_{d,-k}},$$

$$\frac{\mathrm{dlog}\, Z_k}{\mathrm{d}\mu^g_{n,-k}} = \frac{1}{Z_k} \frac{\mathrm{d}Z_k}{\mathrm{d}\mu^g_{n,-k}},$$

$$\frac{\mathrm{d}^2\mathrm{log}\, Z_k}{\mathrm{d}\mu^z_{d,-k}{}^2} = -\frac{1}{Z_k^2} \left(\frac{\mathrm{d}Z_k}{\mathrm{d}\mu^z_{d,-k}}\right)^2 + \frac{1}{Z_k} \frac{\mathrm{d}^2 Z_k}{\mathrm{d}\mu^z_{d,-k}{}^2},$$

$$\frac{\mathrm{d}^2\mathrm{log}\, Z_k}{\mathrm{d}\mu^g_{n,-k}{}^2} = -\frac{1}{Z_k^2} \left(\frac{\mathrm{d}Z_k}{\mathrm{d}\mu^g_{n,-k}}\right)^2 + \frac{1}{Z_k} \frac{\mathrm{d}^2 Z_k}{\mathrm{d}\mu^g_{n,-k}{}^2}.$$

We use these derivatives to update the site parameters in Eq. (15), whilst also converting them back to the precision-adjusted (natural) parameter space, via the following mapping: letting $b_{d,k} = \frac{\mathrm{dlog}\, Z_k}{\mathrm{d}\mu^z_{d,-k}}$ and $c_{d,k} = \frac{\mathrm{d}^2\mathrm{log}\, Z_k}{\mathrm{d}\mu^z_{d,-k}{}^2}$ and for damping parameter ρ,

$$\tau^z_{d,k} = (1-\rho)\tau^z_{d,k} + \frac{\rho}{\eta}\left(\frac{-c_{d,k}}{1 + \zeta_{d,-k}c_{d,k}}\right),$$

$$\nu^z_{d,k} = (1-\rho)\nu^z_{d,k} + \frac{\rho}{\eta}\left(\frac{b_{d,k} - \mu_{d,-k}c_{d,k}}{1 + \zeta_{d,-k}c_{d,k}}\right).$$

Mapping to the natural parameter space in this way makes the updates in the EP algorithm more straightforward (see Alg. 1). The updates for $\tau^g_{n,k}$ and $\nu^g_{n,k}$ are carried out similarly using the derivatives with respect to $\mu^g_{n,-k}$.

C. List of Example Audio Samples Used in Experiments

Table 2. Audio samples used in the experiments. Recordings are mono, 16kHz sample rate.

#	Sample description	Length [ms]
01	speech 1 - male	723
02	speech 2 - male	595
03	speech 3 - male	941
04	speech 4 - male	854
05	speech 5 - male	514
06	speech 6 - female	769
07	speech 7 - female	731
08	speech 8 - female	959
09	speech 9 - female	1270
10	speech 0 - female	1455
11	bamboo flute	2980
12	cello	3300
13	clarinet	2020
14	flute	2000
15	guitar	2652
16	ocarina	1366
17	piano	1846
18	piccolo	1773
19	saxophone	2296
20	toy accordion	1671
21	piano note C	2000
22	piano note E	2000
23	piano note G	2000
24	piano chord mixture	6000
25	piano ground truth C	6000
26	piano ground truth E	6000
27	piano ground truth G	6000