

A Weakly Informative Prior for Resonance Frequencies [†]

Marnix Van Soom *  and Bart de Boer

AI Lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium; bart@ai.vub.ac.be

* Correspondence: marnix@ai.vub.ac.be

† Presented at the 40th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, online, 4–9 July 2021.

Abstract: We derive a weakly informative prior for a set of ordered resonance frequencies from Jaynes' principle of maximum entropy. The prior facilitates model selection problems in which both the number and the values of the resonance frequencies are unknown. It encodes a weakly inductive bias, provides a reasonable density everywhere, is easily parametrizable, and is easy to sample. We hope that this prior can enable the use of robust evidence-based methods for a new class of problems, even in the presence of multiplets of arbitrary order.

Keywords: weakly uninformative prior; resonance frequency; model selection; maximum entropy

1. Introduction

An important problem in the natural sciences is the accurate measurement of resonance frequencies. The problem can be formalized by the following probabilistic model:

$$p(D, x|I) = p(D|x)p(x|I) \equiv \mathcal{L}(x)\pi(x), \quad (1)$$

where D is the data, $x = \{x_k\}_{k=1}^K$ are the K resonance frequencies of interest, and I is the prior information about x . As an example instance of (1), we refer to the vocal tract resonance (VTR) problem discussed in Section 5 for which D is audio recorded from the mouth of a speaker; x are a set of K VTR frequencies, and the underlying model is a sinusoidal regression model. Any realistic problem will include additional model parameters θ , but these have been silently ignored by formally integrating them out of (1), i.e., $p(D, x|I) = \int d\theta p(D, x, \theta|I)$.

In this paper, we assume that the likelihood $\mathcal{L}(x) \equiv p(D|x)$ is given, and our task is to choose an *uninformative* prior $\pi(x) \equiv p(x|I)$ from *limited* prior information I . A conflict arises, however:

The uninformative priors π most commonly chosen to express limited prior information I are, in practice, often precluded by that same I . (2)

The goal of this paper is to describe this conflict (2) and to show how it can be resolved by adopting a specific choice for π . This allows robust inference of the number of resonances K in the important case of such limited prior information I , which in turn enables accurate measurement of the resonance frequencies x with standard methods such as nested sampling [1] or reversible jump MCMC [2].

2. Notation

The symbol π is intended to convey a vague notion of a generally uninformative or weakly informative prior. Definite choices for π are indicated with the subscript i :

$$\pi_i(x) \equiv p(x|\beta_i, I_i), \quad (i = 1, 2, 3), \quad (3)$$



Citation: Van Soom, M.; de Boer, B. A Weakly Informative Prior for Resonance Frequencies. *Phys. Sci. Forum* **2021**, *3*, 2. <https://doi.org/10.3390/psf2021003002>

Academic Editors: Wolfgang von der Linden and Sascha Ranftl

Published: 4 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

where β_i is a placeholder for the hyperparameter specific to π_i . Note that in the plots below and for the experiments in Section 5, the values of the β_i are always set according to Table 1.

Table 1. The values of the hyperparameters β_i used throughout the paper. All quantities are given in units of Hz.

$k \rightarrow$	0	1	2	3	4	5	6	7	8	9	10
$a = \{a_k\}$		200	600	1400	2900	3500					
$b = \{b_k\}$		1100	3500	4000	4500	5500					
$\bar{x}_0 = \{\bar{x}_k\}$	200	500	1000	1500	2000	2500	3000	3500	4000	4500	5000
other	$x_0 = 200$					$x_{\max} = 5500$					

Each π_i uniquely determines a number of important high-level quantities since the likelihood $\mathcal{L}(\mathbf{x})$ and data D are assumed to be given. These quantities are the *evidence* for the model with K resonances:

$$Z_i(K) = \int d^K \mathbf{x} \mathcal{L}(\mathbf{x}) \pi_i(\mathbf{x}), \quad (4)$$

the *posterior*:

$$P_i(\mathbf{x}) = \frac{\mathcal{L}(\mathbf{x}) \pi_i(\mathbf{x})}{Z_i(K)}, \quad (5)$$

and the *information*:

$$H_i(K) = \int d^K \mathbf{x} P_i(\mathbf{x}) \log \frac{P_i(\mathbf{x})}{\pi_i(\mathbf{x})}, \quad (6)$$

which measures the amount of information obtained by updating from prior π_i to posterior P_i , i.e., $H_i(K) \equiv D_{\text{KL}}(P_i | \pi_i)$, where D_{KL} is the Kullback–Leibler divergence.

3. Conflict

The *uninformative priors* π referenced in (2) are of the independent and identically distributed type:

$$\pi(\mathbf{x}) = \prod_{k=1}^K g(x_k | \beta), \quad (7)$$

where $g(x | \beta)$ is any wide distribution with hyperparameters β . A typical choice for g is the uniform distribution over the full frequency bandwidth; other examples include diffuse Gaussians or Jeffreys priors [3–9].

Second, the *limited prior information* I in (2) about K implies that the problem will involve model selection, since each value of K implicitly corresponds to a different model for the data. It is, thus, necessary to evaluate and compare evidence $Z(K) = \int d^K \mathbf{x} \mathcal{L}(\mathbf{x}) \pi(\mathbf{x})$ for each plausible K .

The conflict between these two elements is due to the *label switching problem*, which is a well-known issue in mixture modeling, e.g., [10]. The likelihood functions $\mathcal{L}(\mathbf{x})$ used in models parametrized by resonance frequencies are typically invariant to switching the label k ; i.e., the index k of the frequency x_k has no distinguishable meaning in the model underlying the data. The posterior $P(\mathbf{x}) \propto \mathcal{L}(\mathbf{x}) \pi(\mathbf{x})$ will inherit this exchange symmetry if the prior is of type (7). Thus, if the model parameters \mathbf{x} are well determined by the data D , the posterior landscape will consist of one *primary* mode, which is defined as a mode living in the *ordered region*:

$$\mathcal{R}_K(x_0) = \{\mathbf{x} | x_0 \leq x_1 \leq x_2 \leq \dots \leq x_K\} \quad \text{with} \quad x_0 > 0, \quad (8)$$

and $(K! - 1)$ induced modes, which are identical to the primary mode up to a permutation of the labels k and, thus, live outside of the region $\mathcal{R}_K(x_0)$. The trouble is that correctly taking into account these induced modes during the evaluation of $Z(K)$ requires a surpris-

ing amount of extra work in addition to tuning the MCMC method of choice, and that is the label switching problem in our setting. In fact, there is currently no widely accepted solution for the label switching problem in the context of mixture models either [11,12]. This is, then, how in (2) uninformative priors π are “precluded” by the limited information I : the latter implies model selection, which in turn implies evaluating $Z(K)$, which is hampered by the label switching problem due to the exchange symmetry of the former. Therefore, it seems better to try to avoid it by encoding our preference for primary modes directly into the prior. This results in abandoning the uninformative prior π in favor of the weakly informative prior π_3 , which is proposed in Section 4 as a solution to the conflict.

We use the VTR problem to briefly illustrate the label switching problem in Figure 1. The likelihood $\mathcal{L}(x)$ is described implicitly in Section 5 and is invariant to switching the labels k because the underlying model function (23) of the regression model is essentially a sum of sinusoids, one for each x_k . As frequencies can be profitably thought of as scale variables ([13], Appendix A), the uninformative prior (7) is represented by

$$\pi_1(x) \equiv p(x|x_0, x_{\max}, I_1) = \prod_{k=1}^K h(x_k|x_0, x_{\max}), \quad (9)$$

where $\beta_1 \equiv (x_0, x_{\max})$ are a common lower and upper bound, and

$$h(x|a, b) = \begin{cases} \frac{1}{\log(b/a)} \frac{1}{x} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \quad \text{with } \begin{matrix} a > 0 \\ b < \infty \end{matrix} \quad (10)$$

is the Jeffreys prior, the conventional uninformative prior for a scale variable [although any prior of the form (7) that is sufficiently uninformative would yield essentially the same results.] We have visualized the posterior landscape $P_1(x)$ in Figure 1 by using the pairwise marginal posteriors $P_1(x_k, x_\ell)$ plotted in blue. Note the exchange symmetry of P_1 , which manifests as an (imperfect) reflection symmetry around the dotted diagonal $x_k = x_\ell$ bordering the ordered region $\mathcal{R}_3(x_0)$. The primary mode can be identified by the black dot; all other modes are induced modes. Integrating all $K!$ modes to obtain $Z(K)$ quickly becomes intractable for $Z \gtrsim 4$.

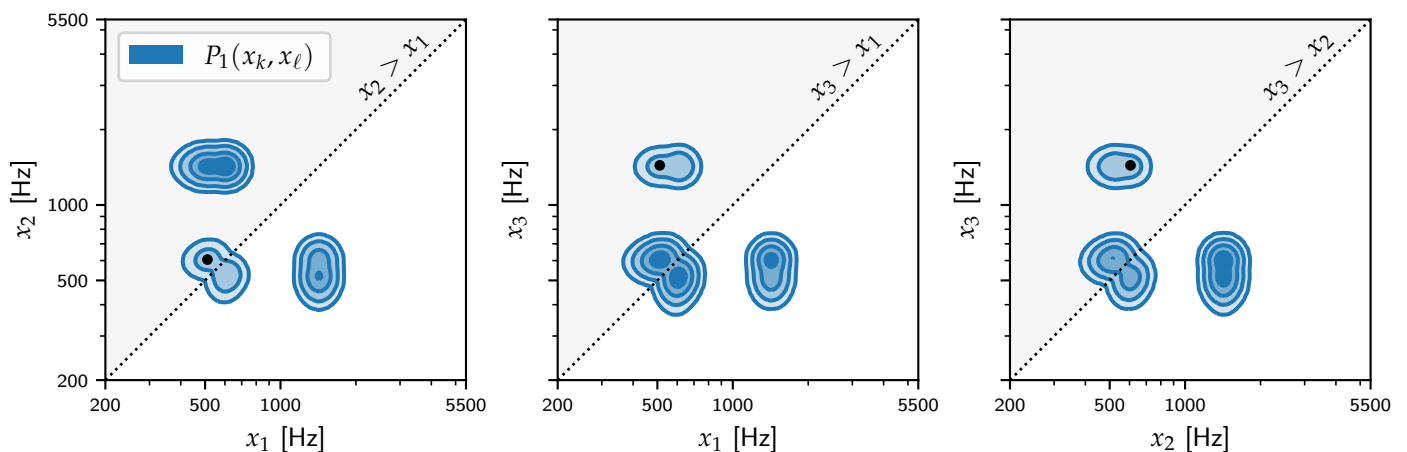


Figure 1. The exchange symmetry of the posterior $P_1(x)$ for a well-determined instance of the VTR problem from Section 5 with $K := 3$. The pairwise marginal posteriors $P_1(x_k, x_\ell)$ are shown using the isocontours of kernel density approximations calculated from posterior samples of x . For each panel, the diagonal $x_k = x_\ell$ is plotted as a dotted line, and the ordered region $\mathcal{R}_3(x_0)$ is shaded in grey. The black dot marks the mean of the primary mode for this problem.

A Simple Way Out?

A simple method out of the conflict is to break the exchange symmetry by assuming specialized bounds for each x_k :

$$\pi_2(x) \equiv p(x|a, b, I_2) = \prod_{k=1}^K h(x_k|a_k, b_k), \quad (11)$$

where $\beta_2 \equiv (a, b)$ with $a = \{a_k\}_{k=1}^K$ and $b = \{b_k\}_{k=1}^K$ being hyperparameters specifying the individual bounds. However, in order to enable the model to detect doublets (a resolved pair of two close frequencies such as the primary mode in the leftmost panel in Figure 1), it is necessary to assign overlapping bounds in (a, b) , presumably by using some heuristic. The necessary degree of overlap increases as the detection of higher order multiplets such as triplets (which can and do occur) is desired, but the more overlap in (a, b) , the more the label switching problem returns. Despite this issue, there will be cases where we have sufficient prior information I to set the (a, b) hyperparameters without too much trouble; the VTR problem is such a case for which the overlapping values of (a, b) up to $K = 5$ are given in Table 1.

4. Solution

Our solution to the conflict (2) is a chain of K coupled Pareto distributions:

$$\pi_3(x) \equiv p(x|\bar{x}_0, I_3) = \prod_{k=1}^K \text{Pareto}(x_k|x_{k-1}, \lambda_k) \quad (12)$$

where

$$\text{Pareto}(x|x_*, \lambda) = \begin{cases} \frac{\lambda x_*^\lambda}{x^{\lambda+1}} & \text{if } x \geq x_* \\ 0 & \text{otherwise} \end{cases} \quad \text{with} \quad \begin{cases} x_* > 0 \\ \lambda > 0, \end{cases} \quad (13)$$

and the hyperparameter $\beta_3 \equiv \bar{x}_0$ is defined as

$$\bar{x}_0 \equiv (\bar{x}_0, \bar{x}), \quad \bar{x}_0 := x_0, \quad \bar{x} = \{\bar{x}_k\}_{k=1}^K, \quad \lambda_k = \frac{\bar{x}_k}{\bar{x}_k - \bar{x}_{k-1}}. \quad (14)$$

From Figure 2, it can be seen that π_3 encodes weakly informative knowledge about K ordered frequencies: (12) and (13) together imply that $\pi_3(x)$ is defined only for $\bar{x} \in \mathcal{R}_K(x_0)$, while nonzero only for $x \in \mathcal{R}_K(x_0)$. In other words, its support is precisely the ordered region $\mathcal{R}_K(x_0)$, which solves the label switching problem underlying the conflict automatically, as the exchange symmetry of π is broken. This is illustrated in Figure 2, where P_3 contracts to a single primary mode, which is just what we would like.

The $K + 1$ hyperparameters \bar{x}_0 in (14) are a common lower bound x_0 plus K expected values of the resonance frequencies \bar{x} . While the former is generally easily determined, the latter may seem difficult to set given the premise of this paper that we dispose only of limited prior information I . Why do we claim that π_3 is only weakly informative if it is parametrized by the expected values of the very things it is supposed to be only weakly informative about? The answer is that for any reasonable amount of data, inference based on π_3 is completely insensitive to the exact values of \bar{x} . Therefore, any reasonable guess for \bar{x}_0 will suffice in practice. For example, for the VTR problem, we simply applied a heuristic where we take $\bar{x}_k = k \times 500$ Hz (see Table 1). This insensitivity is due to the maximum entropy status of π_3 and indicates the weak inductive bias it entails. On a more prosaic level, the heavy tails of the Pareto distributions in (12) ensure that the prior will be eventually overwhelmed by the data no matter how a priori improbable the true value of x is. More prosaic still, in Section 5.1 below we show quantitatively that for the VTR problem π_3 is about as (un)informative as π_2 .

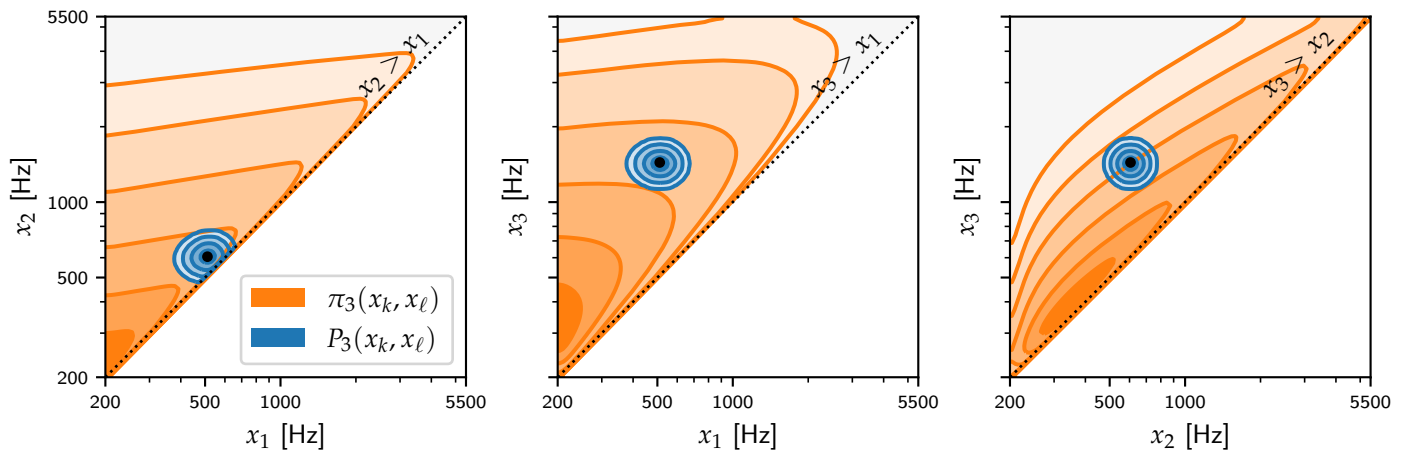


Figure 2. Contraction of prior (π_3) to posterior (P_3) for the application of π_3 to the VTR problem used in Figure 1. The pairwise marginal prior $\pi_3(x_k, x_\ell)$ is obtained by integrating out the third frequency; for example, $\pi_3(x_1, x_2) = \int dx_3 \pi_3(\mathbf{x})$. Unlike P_1 in Figure 1, P_3 exhibits only a single mode that coincides with the primary mode as marked by the black dot.

4.1. Derivation of π_3

Our ansatz consists of interpreting the \mathbf{x} as a set of K ordered scale variables that are bounded from below by x_0 . Starting from (9) and not bothering with the bounds (a, b) , we obtain the improper pdf

$$m(\mathbf{x}) \propto \begin{cases} \prod_{k=1}^K \frac{1}{x_k} & \mathbf{x} \in \mathcal{R}_K(x_0) \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

We can simplify (15) using the one-to-one transformation $\mathbf{x} \leftrightarrow \mathbf{u}$ defined as

$$\begin{aligned} \mathbf{x} \rightarrow \mathbf{u}: \quad u_k &= \log \frac{x_k}{x_{k-1}} \quad (k = 1, 2, \dots, K) \\ \mathbf{u} \rightarrow \mathbf{x}: \quad x_k &= x_0 \exp \sum_{\kappa=1}^k u_\kappa \quad (k = 1, 2, \dots, K) \end{aligned} \quad (16)$$

which yields (with abuse of notation for brevity)

$$m(\mathbf{u}) \propto \begin{cases} 1 & \mathbf{u} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

Since model selection requires proper priors, we need to normalize $m(\mathbf{u})$ by adding extra information (i.e., constraints) to it; we propose to simply fix the K first moments $\langle \mathbf{u} \rangle = \{\langle u_k \rangle\}_{k=1}^K$. This will yield the Pareto chain prior $\pi_3(\mathbf{u})$ directly, expressed in \mathbf{u} space rather than \mathbf{x} space. The expression for $\pi_3(\mathbf{u})$ is found by minimizing the Kullback–Leibler divergence [14]

$$D_{\text{KL}}(\pi_3|m) = \int d^K \mathbf{u} \pi_3(\mathbf{u}) \log \frac{\pi_3(\mathbf{u})}{m(\mathbf{u})}, \quad \text{subject to} \quad \langle \mathbf{u} \rangle \equiv \int d^K \mathbf{u} \mathbf{u} \pi_3(\mathbf{u}) = \bar{\mathbf{u}}, \quad (18)$$

where $\bar{\mathbf{u}} = \{\bar{u}_k\}_{k=1}^K$ are the supplied first moments. This variational problem is equivalent to finding $\pi_3(\mathbf{u})$ by means of Jaynes' principle of maximum entropy with $m(\mathbf{u})$ serving as the invariant measure [15]. Since the exponential distribution $\text{Exp}(x|\lambda)$ is the maximum entropy distribution for a random variable $x \geq 0$ with a fixed first moment $\langle x \rangle = 1/\lambda$, the solution to (18) is

$$\pi_3(\mathbf{u}) = \prod_{k=1}^K \text{Exp}(u_k|\lambda_k), \quad (19)$$

where the rate hyperparameters $\lambda_k = 1/\bar{u}_k$ and

$$\text{Exp}(x|\lambda) = \begin{cases} \lambda \exp\{-\lambda x\} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{with } \lambda > 0. \quad (20)$$

Transforming (19) to x space using (16) finally yields (12), but we still need to express λ_k in terms of \bar{x} —we might find it hard to pick reasonable values of $\bar{u}_k = \overline{\log x_k/x_{k-1}}$ from limited prior information I . For this, we will need the identity

$$\langle x_k \rangle \equiv \int d^K x x_k \pi_3(x) = \frac{\lambda_k}{\lambda_k - 1} \langle x_{k-1} \rangle \quad (k = 1, 2, \dots, K). \quad (21)$$

Constraining $\langle x_k \rangle = \bar{x}_k$ and solving for λ_k , we obtain $\lambda_k = \bar{x}_k / (\bar{x}_k - \bar{x}_{k-1})$, in agreement with (14). Note that the existence of the first marginal moments $\langle x_k \rangle$ requires that $\lambda_k > 1$.

4.2. Sampling from π_3

Sampling from π_3 is trivial because of the independence of the u_k in u space (19). To produce a sample $x' \sim \pi_3(x)$ given the hyperparameter \bar{x}_0 , compute the corresponding rate parameters $\{\lambda_k\}_{k=1}^K$ from (14), and use them in (19) to obtain a sample $u' \sim \pi_3(u)$. The desired x' is then obtained from u' using the transformation (16).

5. Application: The VTR Problem

We now present a relatively simple but realistic instance of the problem of measuring resonance frequencies, which will allow us to illustrate the above ideas. The VTR problem consists of measuring human vocal tract resonance (VTR) frequencies x for each of five representative vowel sounds taken from the CMU ARCTIC database [16]. The VTR frequencies x describe the *vocal tract transfer function* $T(x)$ and are fundamental quantities in acoustic phonetics [17]. The five vowel sounds are recorded utterances of the first vowel in the words $W = \{\text{shore, that, you, little, until}\}$. In order to achieve high-quality VTR frequency estimates \hat{x} , only the quasi-periodic steady-state part of the vowel sound is considered for the measurement. The data D , thus, consists of a string of highly correlated *pitch periods*. See Figure 3 for an illustration of these concepts.

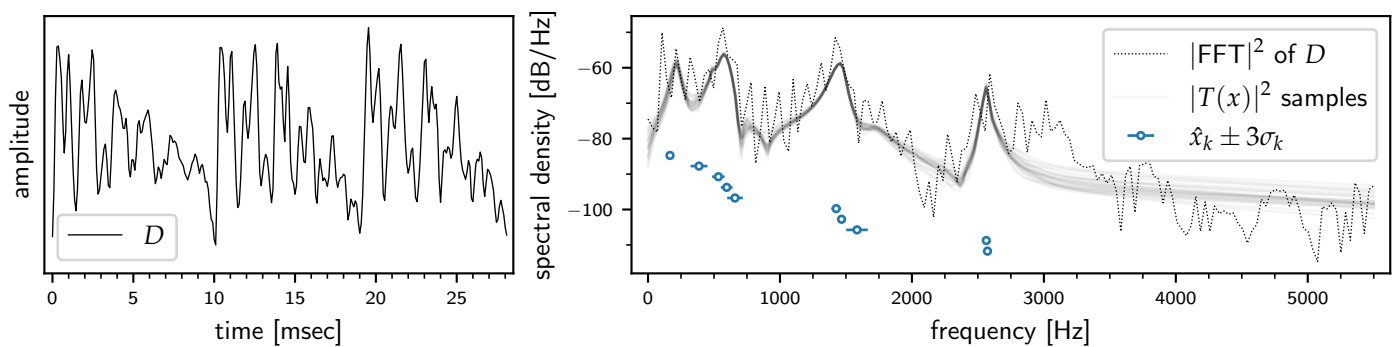


Figure 3. The VTR problem for the case ($D := \text{until}$, $K := 10$). Left panel: The data D , i.e., the quasi-periodic steady-state part, consist of 3 highly correlated pitch periods. Right panel: Inferred VTR frequency estimates $\{\hat{x}_k\}_{k=1}^K$ for $K := 10$ at 3 sigma. They describe the power spectral density of the vocal tract transfer function $|T(x)|^2$, represented here by 25 posterior samples and compared to the Fast Fourier Transform (FFT) of D . All \hat{x}_k are well resolved, and most have error bars too small to be seen on this scale.

The measurement itself is formalized as inference using the probabilistic model (1). The model assumed to underlie the data is the sinusoidal regression model introduced

in [18]; due to limited space, we only describe it implicitly. The sinusoidal regression model assumes that each pitch period $d \in D$ can be modeled as

$$d_t = f(t; A, \alpha, x) + \sigma e_t \quad \text{where} \quad e_t \sim \mathcal{N}(0, 1), \quad (t = 1, 2, \dots, T), \quad (22)$$

where $d = \{d_t\}_{t=1}^T$ is a time series consisting of T samples. The model function

$$f(t; A, \alpha, x) = \sum_{k=1}^K [A_k \cos(x_k t) + A_{K+k} \sin(x_k t)] \exp\{-\alpha_k t\} + \sum_{\ell=1}^L A_{2K+\ell} t^{\ell-1} \quad (23)$$

consists of a sinusoidal part (first Σ) and a polynomial trend correction (second Σ). Note the additional model parameters $\theta = \{A, \alpha, \sigma, L\}$. Formally, given the prior $p(\theta)$ ([18], Section 2.2), the marginal likelihood $\mathcal{L}(x)$ is then obtained as $\mathcal{L}(x) = \int d\theta \mathcal{L}(x, \theta) p(\theta)$, where the complete likelihood $\mathcal{L}(x, \theta)$ is implicitly given by (22) and (23). Practically, we just marginalize out θ from samples obtained from the complete problem $p(D, x, \theta|I)$.

For inference, the computational method of choice is nested sampling [1] using the dynesty library [19–23], which scales roughly as $\mathcal{O}(K^2)$ [24]. Since the VTR problem is quite simple ($H_i(K) \sim 30$ nats), we only perform single nested sampling runs and take the obtained $\log Z_i(K)$ and $H_i(K)$ as point estimates. Full details on the experiments and data are available at <https://github.com/mvsom/frequency-prior>.

5.1. Experiment I: Comparing π_2 and π_3

In Experiment I, we perform a high-level comparison between π_2 and π_3 in terms of *evidence* (4) and *information* (6). The values of the hyperparameters used in the experiment are listed in Table 1. We did not include π_1 in this comparison as the label switching problem prevented convergence of nested sampling runs for $K \geq 4$. The (a, b) bounds for π_2 were based on loosely interpreting the VTRs as formants and consulting formant tables from standard works [25–30]. These allowed us to compile bounds up until the fifth formant such that $K_{\max} = 5$. For π_3 , we simply applied a heuristic where we take $\bar{x}_k = k \times 500$ Hz. We selected x_0 empirically (although a theoretical approach is also possible [31]), and x_{\max} was set to the Nyquist frequency. The role of x_{\max} is to truncate π_3 in order to avoid aliasing effects, since the support of $\pi_3(x_i)$ is unbounded from above. We implemented this by using the following likelihood function in the nested sampling program:

$$\mathcal{L}'(x) = \begin{cases} \mathcal{L}(x) & \text{if } x_k \leq x_{\max} \text{ for all } (k = 1, 2, \dots, K) \\ 0 & \text{otherwise.} \end{cases} \quad (24)$$

First, we compare the influence of π_2 and π_3 on model selection. Given $D \in W$, the posterior probability of the number of resonances K is given by the following.

$$p_i(K) = \frac{Z_i(K)}{\sum_{K'} Z_i(K')} \quad (K = 1, 2, \dots, K_{\max}). \quad (25)$$

The results in the top row of Figure 4a are striking: while $p_2(K)$ shows individual preferences based on D , $p_3(K)$ prefers $K = K_{\max}$ unequivocally.

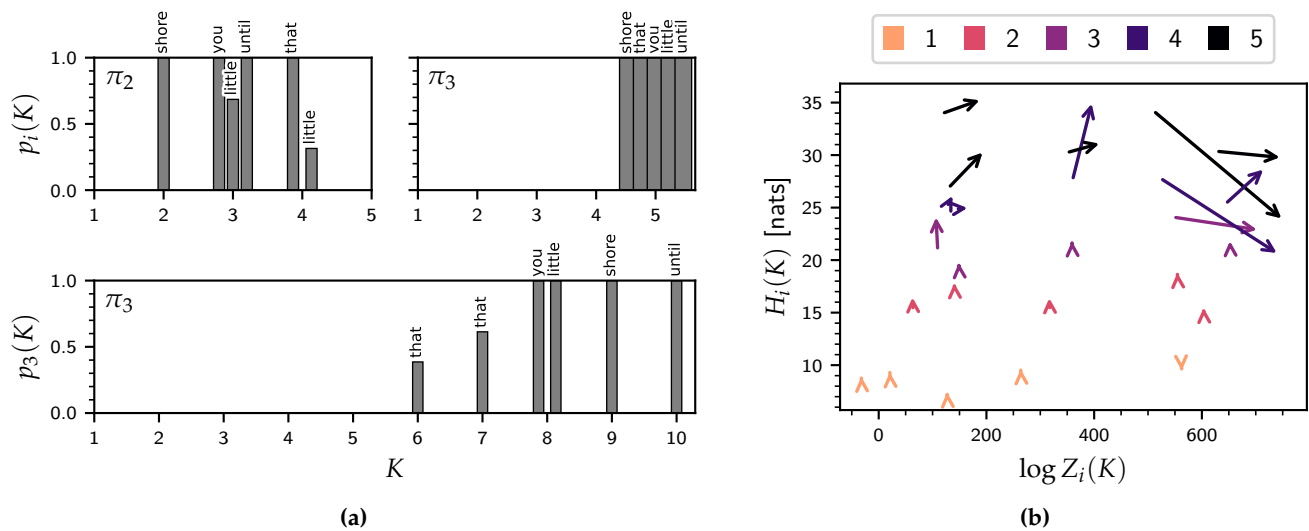


Figure 4. (a) Model selection in Experiment I (top row) and Experiment II (bottom row). (b) In Experiment I, π_2 and π_3 are compared in terms of evidence [$\log Z_i(K)$] and uninformativeness [$H_i(K)$] for each (D, K) . The arrows point from π_2 to π_3 and are color-coded by the value of K . For small values of K , the arrow lengths are too small to be visible on this scale.

Second, in Figure 4b, we compare π_2 and π_3 directly in terms of differences in evidence [$\log Z_i(K)$] and uninformativeness [$H_i(K)$] for each combination (D, K) .

Arrows pointing *eastward* indicate $Z_3(K) > Z_2(K)$. The π_3 prior dominates the π_2 prior in terms of evidence, for almost all values of K , indicating that π_3 places its mass in regions of higher likelihood or, equivalently, that the data were much more probable under π_3 than π_2 . This implies that the hint of π_3 at more structure beyond $K > K_{\max}$ should be taken serious—we investigate this in Section 5.2.

Arrows pointing *northward* indicate $H_3(K) > H_2(K)$, i.e., π_3 is *less* informative than π_2 , since more information is gained by updating from π_3 to P_3 than from π_2 to P_2 . It is observed that π_2 and π_3 are roughly comparable in terms of (un)informativeness.

5.2. Experiment II: ‘Free’ Analysis

We now freely look for more structure in the data by letting K vary up until $K_{\max} = 10$. This goes beyond the capacities of π_1 (because of the label switching problem) and π_2 (because no data are available to set the (a, b) bounds). Thus, the great advantage of π_3 is that we can use a simple heuristic to set \bar{x}_0 and let the model perform the discovering without worrying about convergence issues or the obtained evidence values. The bottom row in Figure 4a shows that model selection for the VTR problem is well-defined, with the most probable values of $K \leq 10$, except for $D = \text{until}$. That case is investigated in Figure 3, where the need for more VTRs (higher K) is apparent from the unmodeled broad peak centered at around 3000 Hz in the FFT power spectrum (right panel). Incidentally, this spectrum also shows that spectral peaks are often resolved into more than one VTR, which underlines the importance of using a prior that enables trouble-free handling of multiplets of arbitrary order. A final observation from the spectrum is the fact that the inferred \hat{x}_k differs substantially from the supplied values in \bar{x} (Table 1), which hints at the weak inductive bias underlying π_3 .

6. Discussion

It is only when the information in the prior is comparable to the information in the data that the prior probability can make any real difference in parameter estimation problems or in model selection problems ([32], p. 9).

Although the weakly informative prior for resonance frequencies π_3 is meant to be overwhelmed, its practical advantage (i.e., solving the label switching problem) will

nonetheless persist, making a real difference in model selection problems even when “the information in the prior” is much smaller than “the information in the data”. In this sense, π_3 is quite unlike the prior referenced in the above quote. Since it will be overwhelmed, all it has to do is provide a reasonable density everywhere (which it does), be easily parametrizable (which it is), and be easy to sample from (which it is).

Thus, we hope that this prior can enable the use of robust evidence-based methods for a new class of problems, even in the presence of multiplets of arbitrary order. The prior is compatible with off-the-shelf exploration algorithms and solves the label switching problem without any special tuning or post processing. It would be interesting to compare it to other approaches, e.g., [33], especially in terms of exploration efficiency. It is valid for any collection of scale variables that is intrinsically ordered, of which frequencies and wavelengths seem to be the most natural examples. Some examples of recent work where the prior could be applied directly are:

- Nuclear magnetic resonance (NMR) spectroscopy [34];
- Resonant ultrasound spectroscopy (a standard method in material science) [35];
- In the analysis of atomic spectra [36], such as X-ray diffraction [37];
- Accurate modeling of instrument noise (in this case LIGO/Virgo noise) [38];
- Model-based Bayesian analysis in acoustics [39].

Author Contributions: Conceptualization, writing, methodology, and analysis: M.V.S. Supervision: B.d.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Flemish AI plan and by the Research Foundation Flanders (FWO) under grant number G015617N.

Acknowledgments: We would like to thank Roxana Radulescu, Timo Verstraeten, and Yannick Jadoul for helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Skilling, J. Nested Sampling for General Bayesian Computation. *Bayesian Anal.* **2006**, *1*, 833–859. [\[CrossRef\]](#)
2. Green, P.J. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* **1995**, *82*, 711–732. [\[CrossRef\]](#)
3. Mark, Y.Z.; Hasegawa-johnson, M. Particle Filtering Approach to Bayesian Formant Tracking. In Proceedings of the IEEE Workshop on Statistical Signal Processing, St. Louis, MO, USA, 28 September–1 October 2003.
4. Zheng, Y.; Hasegawa-Johnson, M. Formant Tracking by Mixture State Particle Filter. In Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, QC, Canada, 17–21 May 2004; Volume 1, pp. 1–565. [\[CrossRef\]](#)
5. Yan, Q.; Vaseghi, S.; Zavarehei, E.; Milner, B.; Darch, J.; White, P.; Andrianakis, I. Formant Tracking Linear Prediction Model Using HMMs and Kalman Filters for Noisy Speech Processing. *Comput. Speech Lang.* **2007**, *21*, 543–561. [\[CrossRef\]](#)
6. Mehta, D.D.; Rudoy, D.; Wolfe, P.J. Kalman-Based Autoregressive Moving Average Modeling and Inference for Formant and Antiformant Tracking. *J. Acoust. Soc. Am.* **2012**, *132*, 1732–1746. [\[CrossRef\]](#)
7. Shi, Y.; Chang, E. Spectrogram-Based Formant Tracking via Particle Filters. In Proceedings of the (ICASSP '03), 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, China, 6–10 April 2003; Volume 1, p. 1. [\[CrossRef\]](#)
8. Deng, L.; Lee, L.J.; Attias, H.; Acero, A. Adaptive Kalman Filtering and Smoothing for Tracking Vocal Tract Resonances Using a Continuous-Valued Hidden Dynamic Model. *IEEE Trans. Audio Speech Lang. Process.* **2007**, *15*, 13–23. [\[CrossRef\]](#)
9. Luberadzka, J.; Kayser, H.; Hohmann, V. Glimpsed Periodicity Features and Recursive Bayesian Estimation for Modeling Attentive Voice Tracking. *Int. Congr. Acoust.* **2019**, *9*, 8.
10. Stephens, M. Dealing with Label Switching in Mixture Models. *J. R. Stat. Soc. Ser. (Stat. Methodol.)* **2000**, *62*, 795–809. [\[CrossRef\]](#)
11. Celeux, G.; Kamary, K.; Malsiner-Walli, G.; Marin, J.M.; Robert, C.P. Computational Solutions for Bayesian Inference in Mixture Models. *arXiv* **2018**, arXiv:1812.07240.
12. Celeux, G.; Fruewirth-Schnatter, S.; Robert, C.P. Model Selection for Mixture Models - Perspectives and Strategies. *arXiv* **2018**, arXiv:1812.09885.
13. Bretthorst, G.L. *Bayesian Spectrum Analysis and Parameter Estimation*; Springer: Berlin/Heidelberg, Germany, 1988.
14. Knuth, K.H.; Skilling, J. Foundations of Inference. *Axioms* **2012**, *1*, 38–73. [\[CrossRef\]](#)
15. Jaynes, E.T. Prior Probabilities. *IEEE Trans. Syst. Sci. Cybern.* **1968**, *4*, 227–241. [\[CrossRef\]](#)
16. Kominek, J.; Black, A.W. The CMU Arctic Speech Databases. In Proceedings of the Fifth ISCA Workshop on Speech Synthesis, Pittsburgh, PA, USA, 14–16 June 2004.

17. Van Soom, M.; de Boer, B. A New Approach to the Formant Measuring Problem. *Proceedings* **2019**, *33*, 29. [[CrossRef](#)]
18. Van Soom, M.; de Boer, B. Detrending the Waveforms of Steady-State Vowels. *Entropy* **2020**, *22*, 331. [[CrossRef](#)] [[PubMed](#)]
19. Speagle, J.S. Dynesty: A Dynamic Nested Sampling Package for Estimating Bayesian Posteriors and Evidences. *arXiv* **2019**, arXiv:1904.02180.
20. Feroz, F.; Hobson, M.P.; Bridges, M. MULTINEST: An Efficient and Robust Bayesian Inference Tool for Cosmology and Particle Physics. *Mon. Not. R. Astron. Soc.* **2009**, *398*, 1601–1614. [[CrossRef](#)]
21. Neal, R.M. Slice Sampling. *Ann. Stat.* **2003**, *31*, 705–767. [[CrossRef](#)]
22. Handley, W.J.; Hobson, M.P.; Lasenby, A.N. POLYCHORD: Nested Sampling for Cosmology. *Mon. Not. R. Astron. Soc.* **2015**, *450*, L61–L65. [[CrossRef](#)]
23. Handley, W.J.; Hobson, M.P.; Lasenby, A.N. POLYCHORD: Next-Generation Nested Sampling. *Mon. Not. R. Astron. Soc.* **2015**, *453*, 4384–4398. [[CrossRef](#)]
24. Buchner, J. Nested Sampling Methods. *arXiv* **2021**, arXiv:2101.09675.
25. Peterson, G.E.; Barney, H.L. Control Methods Used in a Study of the Vowels. *J. Acoust. Soc. Am.* **1952**, *24*, 175–184. [[CrossRef](#)]
26. Hillenbrand, J.; Getty, L.A.; Clark, M.J.; Wheeler, K. Acoustic Characteristics of American English Vowels. *J. Acoust. Soc. Am.* **1995**, *97*, 3099–3111. [[CrossRef](#)]
27. Vallée, N. Systèmes Vocaliques: De La Typologie Aux Prédications, Ph.D. Thesis, Université Stendhal, Grenoble, France, 1994.
28. Kent, R.D.; Vorperian, H.K. Static Measurements of Vowel Formant Frequencies and Bandwidths: A Review. *J. Commun. Disord.* **2018**, *74*, 74–97. [[CrossRef](#)]
29. Vorperian, H.K.; Kent, R.D.; Lee, Y.; Bolt, D.M. Corner Vowels in Males and Females Ages 4 to 20 Years: Fundamental and F1–F4 Formant Frequencies. *J. Acoust. Soc. Am.* **2019**, *146*, 3255–3274. [[CrossRef](#)] [[PubMed](#)]
30. Klatt, D.H. Software for a Cascade/Parallel Formant Synthesizer. *J. Acoust. Soc. Am.* **1980**, *67*, 971–995. [[CrossRef](#)]
31. de Boer, B. Acoustic Tubes with Maximal and Minimal Resonance Frequencies. *J. Acoust. Soc. Am.* **2008**, *123*, 3732. [[CrossRef](#)]
32. Bretthorst, G.L. Bayesian Analysis. II. Signal Detection and Model Selection. *J. Magn. Reson.* **1990**, *88*, 552–570. [[CrossRef](#)]
33. Buscicchio, R.; Roebber, E.; Goldstein, J.M.; Moore, C.J. Label Switching Problem in Bayesian Analysis for Gravitational Wave Astronomy. *Phys. Rev. D* **2019**, *100*, 084041. [[CrossRef](#)]
34. Wilson, A.G.; Wu, Y.; Holland, D.J.; Nowozin, S.; Mantle, M.D.; Gladden, L.F.; Blake, A. Bayesian Inference for NMR Spectroscopy with Applications to Chemical Quantification. *arXiv* **2014**, arXiv:1402.3580.
35. Xu, K.; Marrelec, G.; Bernard, S.; Grimal, Q. Lorentzian-Model-Based Bayesian Analysis for Automated Estimation of Attenuated Resonance Spectrum. *IEEE Trans. Signal Process.* **2019**, *67*, 4–16. [[CrossRef](#)]
36. Trassinelli, M. Bayesian Data Analysis Tools for Atomic Physics. *Nucl. Instruments Methods Phys. Res. Sect. Beam Interact. Mater. Atoms* **2017**, *408*, 301–312. [[CrossRef](#)]
37. Fancher, C.M.; Han, Z.; Levin, I.; Page, K.; Reich, B.J.; Smith, R.C.; Wilson, A.G.; Jones, J.L. Use of Bayesian Inference in Crystallographic Structure Refinement via Full Diffraction Profile Analysis. *Sci. Rep.* **2016**, *6*, 31625. [[CrossRef](#)] [[PubMed](#)]
38. Littenberg, T.B.; Cornish, N.J. Bayesian Inference for Spectral Estimation of Gravitational Wave Detector Noise. *Phys. Rev. D* **2015**, *91*, 084034. [[CrossRef](#)]
39. Xiang, N. Model-Based Bayesian Analysis in Acoustics-A Tutorial. *J. Acoust. Soc. Am.* **2020**. [[CrossRef](#)] [[PubMed](#)]