Response to reviewer 2

Marnix Van Soom Bart de Boer

The paper is very well written and easy to follow. The structure is excellent and the language concise. I find this topic very interesting and the proposed prior seems to produce accurate results in the presented application example. The derivation is free from obvious errors.

We thank the reviewer for the kind words and for reviewing the derivation.

Comments

1. Fig. 1 is hard to understand as P1 and P3 are in some cases overlapping. I would suggest to add an additional figure to separate the display of P1 and P3.

   We agree with the comment and have implemented another factorization of the plots which does not take up too much space (as we were having trouble respecting the page limit). Figure 1 now only shows P_1 and its exchange symmetry. The primary mode has been marked by a black dot. We moved P_3 into Figure 2 to show the contraction from prior to posterior. We made several small changes in the main text to accommodate this change.

   Changes:

   - Figure 1: Show P_1 only, emphasize and mark the mean of the primary mode.
   - Caption of Figure 1: Rewrite first sentence: "The exchange symmetry of the posterior $P_1(\boldsymbol{x})$ for a well-determined instance of the VTR problem from Section **??** with $K := 3$. " and add last "The exchange symmetry of the posterior $P_1(\boldsymbol{x})$ for a well-determined instance of the VTR problem from Section **??** with $K := 3$. "
   - Line 51: Replace last sentence by "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 2: Show π_3 and P_3 on the same plot and mark the mean of the primary mode.
   - Caption of Figure 2: Rewrite first sentence: "Contraction of prior (π_3) to posterior (P_3) during the application of π_3 to the VTR problem used in Figure **??**. and added last: "Unlike P_1 in Figure **??**, P_3 exhibits only a single mode which coincides with the primary mode as marked by the black dot."
   - Line 70: Add extra clarification: "as the exchange symmetry of π is broken."
   - Line 71: Refer to Figure 2: "Figure **??**"

2. Line 114: "The (a, b) bounds for π_2 were based on formant tables from standard works": could you please add a reference and explain why such tables do not exist for $K_{\max} = 10$.

   Formant tables usually only tabulate the first three formants (F1, F2, F3), because it is difficult to measure higher ones (F4, F5) accurately, and because they are more variable among speakers and (thus?) of less interest in the literature. We have changed the wording succinctly to emphasize this and added the references we have used. In addition, we have also added a reference for Table 1 to remind the reader that these formant bounds can be found there.

   Changes in line 114:

- Insert a reference to Table 1 for additional clarity. "The values of the hyperparameters used in the experiment are listed in Table 1."
- Rewrite two sentences about the formant tables. "The $(\boldsymbol{a}, \boldsymbol{b})$ bounds for π_2 were based on loosely interpreting the VTRs as formants and consulting formant tables from standard works [**?**, **?**, **?**, **?**, **?**, **?**]. These allowed us to compile bounds up until the 5th formant, such that $K_{\max} = 5$."

3. In Line 29/30 you mention other possible choices for $g(\boldsymbol{x}|\beta)$: what is the expectation if one uses e.g. a diffuse Gaussian prior in (9)?

As long as the prior is of the $g(\boldsymbol{x}|\beta)$ form (and thus has the exchange symmetry) and sufficiently smooth and wide (what we have called uninformative and "diffuse" in the case of the Gaussian), the results will be essentially identical for any reasonable amount of data. On a higher level, this is because they represent an uninformative state of knowledge and as such will be overwhelmed by the data quickly. We have added a clarification to the main text. We have also made use of the opportunity to add extra references to typical use cases of different kinds of uninformative priors that obey the exchange symmetry.

Changes:

- Lines 47-49: Make it clear that our particular choice for π_1 is largely irrelevant: "[though any prior of the form (**??**) that is sufficiently uninformative would yield essentially the same results.]"
- Line 30: Add references for typical use cases.

4. Concerning model selection: what is the computational complexity to evaluate evidence (4) for each plausible K (approximately)?

The "intrinsic" cost of nested sampling is $\mathcal{O}(NH_i(K))$, where N is the number of live points used and $H_i(K)$ is the information (log compression factor) of the problem. Following (Skilling, 'Nested Sampling's Convergence', 2009), we assume linear scaling of the exploration algorithm and $H_i(K)$ with dimension d. Since $d \propto K$, we arrive roughly at $\mathcal{O}(K^2)$, ignoring the number of live points N. According to Buchner's review of nested sampling (J. Buchner, 'Nested Sampling Methods', arXiv:2101.09675, 2021), this might be a bit optimistic, as empirical evidence points towards cubic scaling in d for some "canonical" problems. But we ignore this refinement as this is not essential for the paper.

Changes on line 112:

- Add complexity: "which scales roughly as $\mathcal{O}(K^2)$ [**?**]"
- Add extra references for the dynesty library

5. Concerning the 'free' analysis: how much does a further increase of $K > 10$ for π_3 improve the output (shown in Fig. 3 (right panel)?

We quickly ran some extra simulations for "until" going up until $K = 16$ to see what would happen. Surprisingly, we found that $p_3(K)$ increases monotonically to $K = 16$, and $p_3(K) \approx 1$ (See Figure 1 below). In other words, it seems that the model still demands further expansion of the vocal tract transfer function $T(x)$, most likely because the polynomial trend correction in Eq. (23) is inadequate (though the model prefers $L = 8$). Looking at the estimated transfer function $T(x)$ for $K = 16$ (Figure 2), the original broad features of $|T(x)|$ at $K = 10$ (showed in Fig. 3 (right panel)) are maintained, but most of the extra 6 VTRs are invested into modeling the fine structure of the broad rightmost peak, as anticipated in the main text ("... 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)", line 137).

But we are inclined to again ignore this refinement to keep the text uncluttered as this is a modeling problem (essentially the polynomial cannot provide enough high frequency content, and this lack of realisticness is being compensated by inserting a lot of detail in the transfer function) and less of an issue of the prior. Indeed, going all the way up to $K = 16$ and still being able to evaluate $Z_3(16)$ would be impossible under π_1 and π_2.

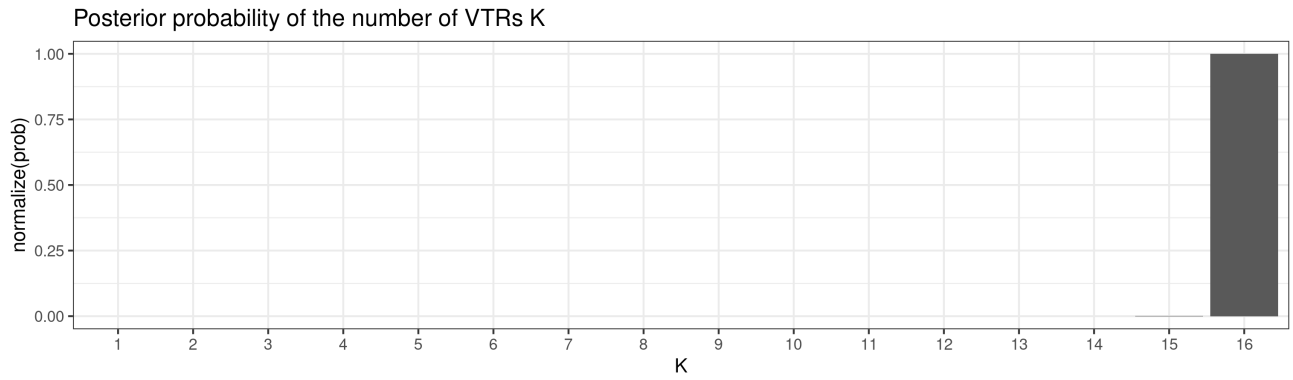

Figure 1: $p_3(K)$

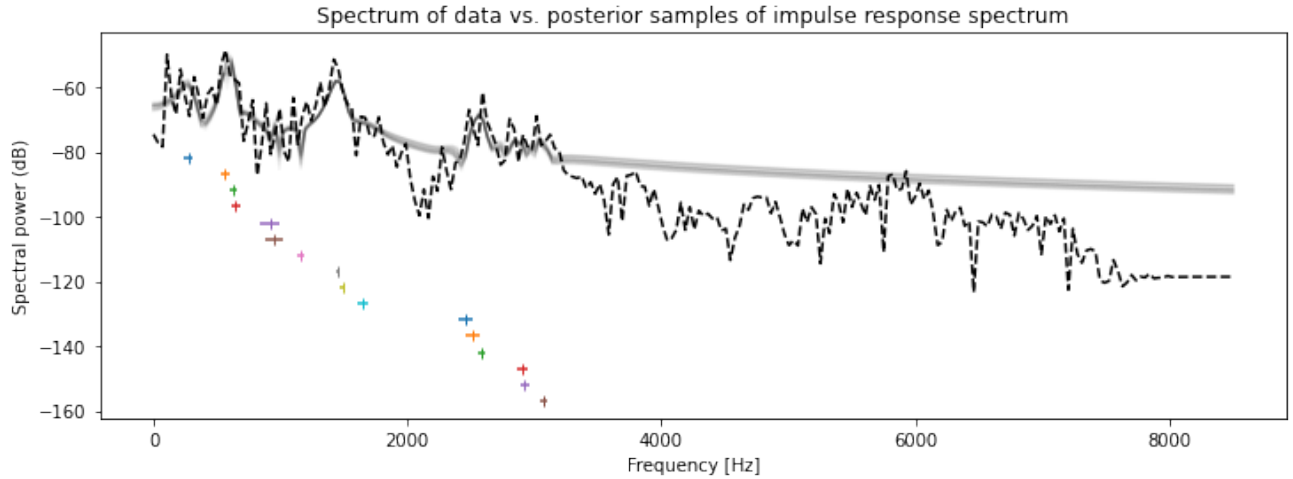

Figure 2: As in Figure 3 (right panel) in the paper but now for $K = 16$ instead of $K = 10$.