Derivation of Analytic solution for the PSTM

Santiago Barreda and T. Florian Jaeger

Introduction

This document will outline the specific implementation of method 6 described in the main paper, along with the derivation of the analytic estimation of the properties of the posterior distribution of ψ_s given the formant pattern and the vowel category. The document below relies on the following notational conventions:

- ψ_s: the speaker-dependent scaling parameter
- $\vec{G}$: the log-transformed formant pattern, i.e. $\vec{G} = \log([F1, F2, F3, \ldots])$
- $\vec{\vec{N}}$: the normalized formant pattern
- $\overrightarrow{\hat{\mu}_{\mathrm{v}}}$: the mean of the normalized formant pattern for a given vowel category
- $\hat{\Sigma}_{\mathrm{v}}$: the covariance matrix of the normalized formant pattern for a given vowel category
- $\hat{\mu}_{\psi_s}$: the mean of the prior distribution of ψ_s
- $\hat{\sigma}_{\psi_s}$: the standard deviation of the prior distribution of ψ_s
- V: the number of vowel categories
- $g0$: the log-transformed $f0$ value
- $\hat{a}_{g0}$: the intercept of the regression line for $f0$ estimation given ψ_s
- $\hat{b}_{g0}$: the slope of the regression line for $f0$ estimation given ψ_s
- $\hat{\sigma}_{g0}$: the standard deviation of the regression of $g0$ estimation given ψ_s
- K: the number of formants
- S: the number of speakers
- $G_{k,\mathrm{v},s}$: the k-th formant value for the s-th speaker in the v-th vowel category
- $N_{k,\mathrm{v},s}$: the normalized k-th formant value for the s-th speaker in the v-th vowel category
- $\hat{\mu}_{k\mathrm{v}}$: the mean of the k-th normalized formant value for the v-th vowel category
- $\hat{\Sigma}_{\mathrm{v},jk}$: the covariance of the j-th and k-th formant values for the v-th vowel category

Parameter Estimation

Here we provide a basic approach to estimation of the relevant parameters. In most cases, more sophisticated techniques may be employed.

ψ_s may be estimated for each speaker as in:

$$\hat{\psi}_s = \frac{1}{(K \cdot V)} \sum_{\mathrm{v}=1}^{V} \sum_{k=1}^{K} G_{k,\mathrm{v},s} \qquad (1)$$

It is very important that the number of tokens be balanced across vowel categories. If a lack of balanced exists, the average of each category can be used in Equation 1 for each speaker. See @b18 for an extended discussion.

Given an estimate of ψ_s, the normalized formant pattern may be estimated as in:

$$\vec{\hat{N}} = \vec{G} - \hat{\psi}_s \qquad (2)$$

The mean of the normalized formant pattern for each vowel category may be estimated using these normalized vectors as in:

$$\vec{\hat{\mu}} = [\hat{\mu}_{1\mathrm{v}}, \hat{\mu}_{2\mathrm{v}}, \hat{\mu}_{3\mathrm{v}}] = \hat{\mu}_{k\mathrm{v}} = \frac{1}{S} \sum_{s=1}^{S} N_{k,\mathrm{v},s} \qquad (3)$$

Again, balance is important. If the number of tokens is not balanced across vowel speakers, the average of each category can be used in Equation 2 for each speaker.

We will discuss estimation of the template covariance matrices in the pooled and category-specific cases. In both situations we assume balance across vowels and listeners, and no missing observations. The more complicated case of covariance estimation in the presence of missing or unbalanced data is left as an exercise for the reader. The covariance matrix of the normalized formant pattern for each vowel category may be estimated as in:

$$\hat{\Sigma}_{\mathrm{v},jk} = \frac{1}{S-1} \sum_{s=1}^{S} \left(N_{j,\mathrm{v},s} - \hat{\mu}_{j,\mathrm{v}}\right) \left(N_{k,\mathrm{v},s} - \hat{\mu}_{k,\mathrm{v}}\right) \qquad (4)$$

And we can estimate the pooled covariance across all categories as in:

$$\hat{\Sigma}_{jk} = \frac{1}{(S \cdot V - 1)} \sum_{\mathrm{v}=1}^{V} \sum_{s=1}^{S} \left(N_{j,\mathrm{v},s} - \hat{\mu}_{j,\mathrm{v}}\right) \left(N_{k,\mathrm{v},s} - \hat{\mu}_{k,\mathrm{v}}\right) \qquad (5)$$

The prior mean of ψ_s can be estimated from any large dataset as in:

$$\hat{\mu}_{\psi_s} = \frac{1}{S} \sum_{s=1}^{S} \hat{\psi}_s \qquad (6)$$

and the prior standard deviation of ψ_s can be estimated as in:

$$\hat{\sigma}_{\psi_s} = \sqrt{\frac{1}{S-1} \sum_{s=1}^{S} \left(\hat{\psi}_s - \hat{\mu}_{\psi_s}\right)^2} \qquad (7)$$

Obviously, the mean and standard deviation estimates will depend heavily on the sorts of speakers included in the data. For estimates that reflect the human population in general, speakers of a wide range of sizes are needed.

The intercept and slope of the regression line for $f0$ estimation given ψ_s can be estimated using the ψ_s estimates for each speaker and the g0 values corresponding to these speakers in the database. For a vector `psi` of ψ_s estimates for the speaker who produced each token, and a vector `g0` of corresponding g0 values for each token, the intercept and slope can be estimated as in:

```
lm (g0 ~ psi)
```

The resulting model intercept, slope, and residual error provide estimates of $\hat{a}_{g0}$, $\hat{b}_{g0}$, and $\hat{\sigma}_{g0}$, respectively.

Derivation of Method 6 Posterior Distributions

According to implementation described in the main article, the posterior probability related to method 6 is equal to the product of the densities in Equation 8.

$$
\begin{aligned}
P(\vec{G}|\mathrm{v},\psi_s) \cdot P(g0|\psi_s) \cdot P(\psi_s) \cdot P(\mathrm{v}) = \\
\mathrm{MVN}(\vec{N} \mid \overrightarrow{\hat{\mu}_{\mathrm{v}}}, \hat{\Sigma}_{\mathrm{v}}) \cdot \\
N(\hat{a}_{g0} + \hat{b}_{g0} \cdot \psi_s, \hat{\sigma}_{g0}) \cdot \\
N(\hat{\mu}_{\psi_s}, \hat{\sigma}_{\psi_s}) \cdot \\
P(\mathrm{v})
\end{aligned}
\tag{8}
$$

These are: the likelihood of the formant pattern given the vowel category and value of ψ_s, the probability of observing a given log-transformed $f0$ (i.e. $g0$) value given ψ_s, the prior distribution of ψ_s, and the prior probability of the vowel. If the prior probability of the vowel is assumed to be equal across all V categories, this can be implemented as in:

$$
\begin{aligned}
P(\vec{G}|\mathrm{v},\psi_s) \cdot P(g0|\psi_s) \cdot P(\psi_s) \cdot P(\mathrm{v}) = \\
\left((2\pi)^{k/2}\left|\widehat{\Sigma_{\mathrm{v}}}\right|^{1/2}\right)^{-1} \exp\left(-1/2 \cdot (\vec{N} - \overrightarrow{\hat{\mu}_{\mathrm{v}}})^T \cdot (\hat{\Sigma}_{\mathrm{v}})^{-1} \cdot (\vec{N} - \overrightarrow{\hat{\mu}_{\mathrm{v}}})\right) \cdot \\
\frac{1}{\hat{\sigma}_{g0}^2\sqrt{2\pi}} \exp\left(-1/2 \cdot (g0 - \hat{a}_{g0} + \hat{b}_{g0} \cdot \psi_s)^2/\hat{\sigma}_{g0}^2\right) \cdot \\
\frac{1}{\hat{\sigma}_{\psi_s}^2\sqrt{2\pi}} \exp\left(-1/2 \cdot (\psi_s - \hat{\mu}_{\psi_s})^2/\hat{\sigma}_{\psi_s}^2\right) \cdot \\
1/V
\end{aligned}
\tag{9}
$$

Where $\vec{N}$ represents the normalized formant pattern. To find the value of ψ_s that maximizes the value of this product, we rely on two useful properties of normal distributions: the product of two normal distributions is a normal distribution, and the logarithm of a normal distribution is a quadratic function. As a result, the derivative of the logarithm of this equation with respect to ψ_s will be a line, and setting this line equal to zero will give us the value of ψ_s that maximizes the posterior probability. We explain this process in detail in the following sections, for now we jump to the result. The derivative (with respect to ψ_s) of the logarithm of Equation 9 is shown in Equation 10.

$$\begin{gathered}\frac{\partial}{\partial \psi_s} \log\left(P\left(\vec{G}|\mathrm{v}, \psi_{s,\mathrm{v}}\right) \cdot P(g0|\psi_s) \cdot P(\psi_s)\right) = \\ \sum\left(\vec{N} \cdot \hat{\Sigma}_{\mathrm{v}}^{-1}\right) - \sum\left(\hat{\Sigma}_{\mathrm{v}}^{-1}\right) \cdot \psi_s \cdot \\ \left(\frac{g0 \cdot \hat{b}_{g0}}{\hat{\sigma}_{g0}^2} - \frac{\hat{a}_{g0} \cdot \hat{b}_{g0}}{\hat{\sigma}_{g0}^2}\right) - \frac{\hat{b}_{g0}^2}{\hat{\sigma}_{g0}^2} \cdot \psi_s \cdot \\ \frac{\hat{\mu}_{\psi_s}}{\hat{\sigma}_{\psi_s}^2} - \frac{1}{\hat{\sigma}_{\psi_s}^2} \cdot \psi_s\end{gathered} \tag{10}$$

Each line above each contains a term that is multiplied by ψ_s and a term that is independent of ψ_s. The terms in Equation 10 can be rearranged to reflect a single line with an intercept equal to $\hat{a}_{\frac{\partial}{\partial \psi_s}}$ and a slope equal to $\hat{b}_{\frac{\partial}{\partial \psi_s}}$, as in Equation 11.

$$\begin{gathered}\frac{\partial}{\partial \psi_s} \log\left(P\left(\vec{G}|\mathrm{v}, \psi_s\right) \cdot P(g0|\psi_s) \cdot P(\psi_s)\right) = \hat{a}_{\frac{\partial}{\partial \psi_s}} + \hat{b}_{\frac{\partial}{\partial \psi_s}} \cdot \psi_s \\ \text{Where:} \\ \hat{a}_{\frac{\partial}{\partial \psi_s}} = \sum\left(\vec{N} \cdot \hat{\Sigma}_{\mathrm{v}}^{-1}\right) + \left(\frac{g0 \cdot \hat{b}_{g0}}{\hat{\sigma}_{g0}^2} - \frac{\hat{a}_{g0} \cdot \hat{b}_{g0}}{\hat{\sigma}_{g0}^2}\right) + \frac{\hat{\mu}_{\psi_s}}{\hat{\sigma}_{\psi_s}^2} \\ \hat{b}_{\frac{\partial}{\partial \psi_s}} = \sum\left(\hat{\Sigma}_{\mathrm{v}}^{-1}\right) + \frac{\hat{b}_{g0}^2}{\hat{\sigma}_{g0}^2} + \frac{1}{\hat{\sigma}_{\psi_s}^2} \psi_s\end{gathered} \tag{11}$$

Setting $0 = \hat{a}_{\frac{\partial}{\partial \psi_s}} + \hat{b}_{\frac{\partial}{\partial \psi_s}} \cdot \psi_s$ and solving for ψ_s (i.e. $\psi_s = -a/b$) gives the mean of the posterior of ψ_s, and $-1/\sqrt{b}$ gives its standard deviation. This value can then be used in Equation 9 to find the posterior density for that value of ψ_s.

The probability of the formant pattern given of the vowel category and ψ_s

The multivariate normal density of the vowel category given the formant pattern and ψ_s is presented in Equation 11.

$$\mathrm{MVN}\left(\overrightarrow{\hat{\mu}_{\mathrm{v}}}, \hat{\Sigma}^{\mathrm{v}}\right) = \frac{1}{(2\pi)^{k/2}\left|\hat{\Sigma}_{\mathrm{v}}\right|^{1/2}} \exp\left(-1/2 \cdot \left(\vec{N} - \overrightarrow{\hat{\mu}_{\mathrm{v}}}\right)^T \cdot \left(\hat{\Sigma}_{\mathrm{v}}\right)^{-1} \cdot \left(\vec{N} - \overrightarrow{\hat{\mu}_{\mathrm{v}}}\right)\right) \tag{12}$$

The multivariate normal density has only one part that depends on ψ_s; the rest are constants which won't matter once we take the derivative. The part that depends on ψ_s is shown in Equation 13.

$$-1/2 \cdot \left(\vec{N} - \overrightarrow{\hat{\mu}_{\mathrm{v}}}\right)^{T} \cdot \left(\hat{\Sigma}_{\mathrm{v}}\right)^{-1} \cdot \left(\vec{N} - \overrightarrow{\hat{\mu}_{\mathrm{v}}}\right) \qquad (13)$$

Since $\vec{N} = \vec{G} - \overrightarrow{\psi_s}$ in the log-mean framework, we can rewrite Equation 13 as in:.

$$-1/2 \cdot \left(\vec{G} - \overrightarrow{\psi_s} - \overrightarrow{\hat{\mu}_{\mathrm{v}}}\right)^{T} \cdot \hat{\Sigma}_{\mathrm{v}}^{-1} \cdot \left(\vec{G} - \overrightarrow{\psi_s} - \overrightarrow{\hat{\mu}_{\mathrm{v}}}\right) \qquad (14)$$

To get the derivative of Equation 14 with respect to ψ_s we break it up into inner and outer functions, i.e. $h\big(g(x)\big)$, and apply the chain rule. We define the outer ($h(x)$) and inner ($g(\psi_s)$) functions as in Equation 15 and Equation 16.

$$h(x) = -1/2 \cdot x^{\mathrm{T}} \cdot \hat{\Sigma}_{\mathrm{v}}^{-1} \cdot x \qquad (15)$$

$$g(\psi_s) = x = \vec{G} - \overrightarrow{\psi_s} - \overrightarrow{\hat{\mu}_{\mathrm{v}}} \qquad (16)$$

The derivative of $h(x)$ with respect to x is given in Equation 17, and the derivative of $g(\psi_s)$ with respect to ψ_s is given in Equation 18.

$$h'(x) = -x \cdot \hat{\Sigma}_{\mathrm{v}}^{-1} \qquad (17)$$

$$g'(\psi_s) = -I = [-1,\ -1,\ -1] \qquad (18)$$

Completing the chain rule, we multiply the derivative of the outer function $h(\cdot)$ and the inner function $g(\cdot)$ as shown in Equation 19.

$$h'(x) \cdot g'(\psi_s) = \left(-x \cdot \hat{\Sigma}_{\mathrm{v}}^{-1}\right) \cdot -I \qquad (19)$$

The product of these is shown in Equation 20, and due to the properties of matrix multiplication (since $I = [1,\ 1,\ 1]$), this reduces as in Equation 21.

$$-\left(\vec{G} - \overrightarrow{\psi_s} - \overrightarrow{\hat{\mu}_{\mathrm{v}}}\right) \cdot \hat{\Sigma}_{\mathrm{v}}^{-1} \cdot -I \qquad (20)$$

$$\sum\left(\left(\vec{G} - \overrightarrow{\psi_s} - \overrightarrow{\hat{\mu}_{\mathrm{v}}}\right) \cdot \hat{\Sigma}_{\mathrm{v}}^{-1}\right) \qquad (21)$$

Which we can rearrange to arrive at the the organization presented in **?@eq-03**.

$$\begin{aligned}
\sum\left(\left(\vec{G} - \overrightarrow{\hat{\mu}_{\mathrm{v}}}\right) \cdot \hat{\Sigma}_{\mathrm{v}}^{-1}\right) - \left(\psi_s \cdot \hat{\Sigma}_{\mathrm{v}}^{-1}\right) &= \\
\sum\left(\left(\vec{G} - \overrightarrow{\hat{\mu}_{\mathrm{v}}}\right) \cdot \hat{\Sigma}_{\mathrm{v}}^{-1}\right) - \psi_s \cdot \left([1,1,1] \cdot \hat{\Sigma}_{\mathrm{v}}^{-1}\right) &= \\
\sum\left(\left(\vec{G} - \overrightarrow{\hat{\mu}_{\mathrm{v}}}\right) \cdot \hat{\Sigma}_{\mathrm{v}}^{-1}\right) - \psi_s \cdot \sum\left(\hat{\Sigma}_{\mathrm{v}}^{-1}\right)&
\end{aligned}$$

The prior distribution of ψ_s

The log density of the normal distribution of ψ_s is shown in Equation 22.

$$\log\left(N\left(\psi_s | \hat{\mu}_{\psi_s}, \hat{\sigma}_{\psi_s}\right)\right) = \log\left(\frac{1}{\hat{\sigma}_{\psi_s}^2\sqrt{2\pi}} \exp\left(-1/2 \cdot \left(\psi_s - \hat{\mu}_{\psi_s}\right)^2 / \hat{\sigma}_{\psi_s}^2\right)\right) \qquad (22)$$

This has only one part that depends on ψ_s, isolated and expanded in Equation 23.

$$-1/2 \cdot \left(\psi_s - \hat{\mu}_{\psi_s}\right)^2 / \hat{\sigma}_{\psi_s}^2 = -\left(\psi_s^2 - 2\psi_s\hat{\mu}_{\psi_s} + \hat{\mu}_{\psi_s}^2\right) / \left(2\hat{\sigma}_{\psi_s}^2\right) \qquad (23)$$

The derivative of this with respect to ψ_s is a line with a slope and an intercept, as shown in Equation 24.

$$-\left(2\psi_s - 2\hat{\mu}_{\psi_s}\right)/2\hat{\sigma}_{\psi_s}^2 = \left(\hat{\mu}_{\psi_s} - \psi_s\right)/\hat{\sigma}_{\psi_s}^2 = \frac{\hat{\mu}_{\psi_s}}{\hat{\sigma}_{\psi_s}^2} - \frac{1}{\hat{\sigma}_{\psi_s}^2}\psi_s \qquad (24)$$

The probability of $g0$ given ψ_s

The log density of $N\left(g0|\hat{\mu}_{g0}, \hat{\sigma}_{g0}^2\right)$ has only one part that depends on ψ_s (the rest are constants which won't matter once we take the derivative). This is presented in Equation 25.

$$-1/2 \cdot \left(g0 - \hat{\mu}_{g0}\right)^2 / \hat{\sigma}_{g0}^2 \qquad (25)$$

Since for us $\mu_{g0} = \hat{a}_{g0} + \hat{b}_{g0} \cdot \psi_s$, we can rewrite Equation 24 as shown in Equation 26.

$$-1/2 \cdot \left(g0 - \left(\hat{a}_{g0} + \hat{b}_{g0} \cdot \psi_s\right)\right)^2 / \hat{\sigma}_{g0}^2 \qquad (26)$$

We can use the chain rule to find the derivative. The outer and inner functions are shown in Equation 27.

$$\begin{aligned} h(x) &= -1/2 \cdot (g0 - x)^2 / \hat{\sigma}_{g0}^2 \\ g(\psi_s) &= \hat{a}_{g0} + \hat{b}_{g0} \cdot \psi_s \end{aligned} \qquad (27)$$

The derivative of $h(x)$ with respect to x is shown in Equation 28, which can be presented as in Equation 29.

$$h'(x) = (g0 - x)/\hat{\sigma}_{g0}^2 = \left(g0 - \left(\hat{a}_{g0} + \hat{b}_{g0} \cdot \psi_s\right)\right)/\hat{\sigma}_{g0}^2 \qquad (28)$$

$$h'(x) = \frac{g0}{\hat{\sigma}_{g0}^2} - \frac{\hat{a}_{g0}}{\hat{\sigma}_{g0}^2} - \frac{\hat{b}_{g0}}{\hat{\sigma}_{g0}^2} \cdot \psi_s \qquad (29)$$

The derivative of $g(\psi_s)$ with respect to ψ_s is shown in Equation 30.

$$g'(\psi_s) = \hat{b}_{g0} \qquad (30)$$

We multiply the derivative of the outer function h and the inner function g to complete the chain rule, as in Equation 31.

$$h'\left(g(x)\right) \cdot g'(x) = \left(\frac{g0}{\hat{\sigma}_{g0}^2} - \frac{\hat{a}_{g0}}{\hat{\sigma}_{g0}^2} - \frac{\hat{b}_{g0}}{\hat{\sigma}_{g0}^2} \cdot \psi_s\right) \cdot \left(\hat{b}_{g0}\right) \qquad (31)$$

The elements composing the intercept, which do not multiply by ψ_s, are grouped in parentheses. This results in the organization shown in Equation 32.

$$\left(\frac{g0 \cdot \hat{b}_{g0}}{\hat{\sigma}_{g0}^2} - \frac{\hat{a}_{g0} \cdot \hat{b}_{g0}}{\hat{\sigma}_{g0}^2}\right) - \frac{\hat{b}_{g0}^2}{\hat{\sigma}_{g0}^2} \cdot \psi_s \qquad (32)$$

Calculating the integral of the likelihood function and posterior distribution

To find the integral of the posterior distribution (or the likelihood function), we rely on the relationship between the standard deviation, the peak density, and the integral of Gaussian-shaped functions. The peak density value of a Gaussian-shaped function, i.e. where $x = \mu$ is given in Equation 33.

$$\mathcal{D}_{N(\cdot)} = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \qquad (33)$$

The integral of a Gaussian probability distribution along the number line will always be 1. However, Gaussian-shaped likelihood functions can have integrals much smaller than or greater than 1. When this occurs, the the maximum density along this curve ($\mathcal{D}_{f(\cdot)}$) can be very different from the peak density of a Gaussian probability distribution ($\mathcal{D}_{N(\cdot)}$) with the same mean and variance. For example, if the peak density of the likelihood function is twice as high as the Gaussian distribution with the same mean and standard deviation, then we know that the likelihood function must have an integral twice as large as the equivalent probability distribution. Since probability distributions have integrals equal to one, this means that we can calculate the integral of Gaussian-shaped likelihoods and posterior distributions ($f(\cdot)$) as in Equation 34.

$$\int_{-\infty}^{\infty} f(x)dx = \frac{\mathcal{D}_{f(x)}}{\mathcal{D}_{N(x)}} = \mathcal{D}_{f(x)} \cdot \sqrt{2\pi\hat{\sigma}^2} \qquad (34)$$