Supplementary materials

1 Sub-model for the categorization of spectral characteristics into phonemes

The aim of this model is to derive the probability distributions $P(\Phi \mid S)$ that were introduced in the main article.

1.0.1 Description

Variables Φ and S are the only variables involved in this sub-model. They are defined according to the main text as:

- **Φ** represents phonemes: $\Phi = \{$/i/, /e/, /ɛ/, /a/, /ɑ/, /ɔ/, /k/, /00/$\}$.
- **S** represents the spectral characteristics of the acoustic signal: $S = (F_1, F_2, F_3)$.

Decomposition The joint probability distribution $P(\Phi\ S)$ is decomposed as

$$P(\Phi\ S) = P(\Phi)\ P(S \mid \Phi). \tag{1}$$

Parametric forms

$P(\Phi)$ corresponds to the knowledge that we have *a priori* about phonemes Φ. Even if this knowledge is certainly language dependent, GEPPETO does not introduce any *a priori* in this sense. $P(\Phi)$ is therefore assumed to be a uniform probability distribution:

$$P(\Phi) = \frac{1}{8}. \tag{2}$$

$P(S \mid \Phi)$ corresponds to the knowledge about the spectral characteristics of the acoustic signal produced for each phoneme Φ. The "no-phoneme" category /00/ was essentially introduced for a better discrimination of phonemes. The probability distribution corresponding to this category is chosen to be uniform over the domain D_S of spectral characteristics. Denoting by $|D_S|$ the volume of this domain, we define:

$$P(S \mid [\Phi = /00/]) = \begin{cases} \frac{1}{|D_S|} & \text{if } S \in D_S \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

For all other values of Φ, we define the corresponding probability distributions $P(S \mid \Phi)$, based on the elliptic regions characterizing each phoneme in GEPPETO. A natural choice are Gaussian probability distributions defined in terms of the corresponding ellipsoids regions and truncated to the range of the domain D_S:

$$P(S \mid [\Phi = \phi]) \quad = \quad \begin{cases} \frac{1}{Z_\phi(\kappa_S)} \mathcal{G}(S \;;\; \mu_\phi,\; \kappa_S^2 \Gamma_\phi) & \text{if } S \in D_S \\ 0 & \text{otherwise} \end{cases} \tag{4}$$
$$\text{for } \phi \in \{ \text{ /i/, /e/, /ɛ/, /a/, /ɑ/, /ɔ/, /k/ } \}$$

μ_ϕ is the center of the ellipsoid attributed to phoneme ϕ, Γ_ϕ is the symmetric matrix defining the quadratic form characterizing the corresponding ellipsoid and $Z_\phi(\kappa_S)$ is the normalization term due to truncation to the domain D_S. The parameter κ_S is introduced in order to modulate the dispersion of the probability distribution.

1.0.2 Question

The goal of the model is to compute the probability $P(\Phi \mid S)$ giving the confidence on categorizing the given spectral characteristics S into phoneme Φ. This is expressed as

$$P(\Phi \mid S) = \frac{P(S\ \Phi)}{P(S)}, \tag{5}$$

with

$$P(S) = \sum_{\Phi} P(S\ \Phi). \tag{6}$$

Making use of the decomposition given by Equation (1), this gives:

$$\begin{aligned} P(\Phi \mid S) &= \frac{P(\Phi)P(S \mid \Phi)}{\sum\limits_{\phi} P(\Phi)P(S \mid \Phi)} \\ &= \frac{P(S \mid \Phi)}{\sum\limits_{\phi} P(S \mid \Phi)}, \end{aligned} \tag{7}$$

where last line made use of the fact that $P(\Phi)$ is constant.

With equation (7) we have completely specified the desired probability distribution $P(\Phi \mid S)$. This ends the definition of this sub-model.

2 Equivalence of models

The probability distribution $P(M^{1:3} \mid \Phi^{1:3}\ [C_m = L])$ characterizes the set of every sequence of control variables $M^{1:3}$ with its probability to achieve the desired sequence of phoneme $\Phi^{1:3}$ with the "minimum effort" constraint ($C_m = L$). If we look for the most probable solutions, the Bayesian model become equivalent to the optimal control approach as we will now see. This equivalence comes from the fact that the cost function optimized by GEPPETO is proportional to the negative log probability of the inferred control variables. Thus, finding the more probable control parameters under the Bayesian model is equivalent to minimizing the cost function defined by GEPPETO. For simplicity we will derive this proof for a sequence of two phonemes, the result for a sequence of three phonemes being easily obtained in the same way.

For a sequence of two phonemes we have

$$P(M^{1:2} \mid \Phi^{1:2}\ [C_m = L]) \propto P(\Phi^1 \mid S^*(M^1))\ P(\Phi^2 \mid S^*(M^2))\ e^{-|M^2-M^1|}, \tag{8}$$

where the parameter κ_M has been set to 1. Let us rewrite Equation (8) in the following form

$$P(M^{1:2} \mid \Phi^{1:2}\ [C_m = L]) \quad \propto \quad e^{-\mathcal{L}_B} \tag{9}$$

where

$$\mathcal{L}_B \quad = \quad -\ln P(\Phi^1 \mid S^*(M^1)) - \ln P(\Phi^2 \mid S^*(M^2)) + |M^2 - M^1|. \tag{10}$$

As will appears shortly $\mathcal{L}_B$ can bee seen as a Lagrangian associated to the Bayesian model. We now unpack the form of this expression in order to compare it to GEPPETO.

It can already be noted that the last term in Equation (10) corresponds to the cost function of GEPPETO, as it is just the Euclidean distance between the two motor control variables M^1 and M^2.

Since the form of the distributions $P(\Phi \mid S)$ are close to step functions on the elliptic domains defined for each phoneme (Figure 4 and 5 of the main text), we approximate them as:

$$P(\Phi^i \mid S^*(M^i)) \quad = \quad \begin{cases} 1 & \text{if } S^*(M^i) \in \mathcal{E}_{\Phi^i} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where $S^*(M^i)$ stands for the spectral characteristics of the acoustic signal corresponding to M^i and $\mathcal{E}_\Phi^i$ correspond to the elliptic domain characterizing Φ^i. Therefore

$$-\ln P(\Phi^i \mid S^*(M^i)) \quad = \quad \begin{cases} 0 & \text{if } S^*(M^i) \in \mathcal{E}_{\Phi^i} \\ \infty & \text{otherwise} \end{cases} \tag{12}$$

On the other hand, the optimization algorithm in GEPPETO minimizes the cost function

$$\mathcal{F}_c(M^{1:2}) \quad = \quad |M^2 - M^1| \tag{13}$$

under the perceptual constraint

$$\mathcal{A}_c(M^i, \Phi^i) \quad = \quad \begin{cases} 0 & \text{if } S^*(M) \in \mathcal{E}_{\Phi^i} \\ \infty & \text{otherwise} \end{cases} \tag{14}$$

for $i \in \{1, 2\}$. We already note that the form of these constraints are identical to the one approximated in Equation (12).

Optimization under constraints is performed in GEPPETO by gradient descent on the Lagrangian defined by

$$\mathcal{L}_G \quad = \quad \mathcal{F}_c + \mathcal{A}_c^1 + \mathcal{A}_c^2, \tag{15}$$

Hence, it appears that $\mathcal{L}_G$ is equal to $\mathcal{L}_B$ in equation (10) and therefore

$$P(M^{1:2} \mid \Phi^{1:2}\ [C_m = L]) \quad \propto \quad e^{-\mathcal{L}_G}. \tag{16}$$

This shows that finding a sequence of control variables $M^{1:2}$ that maximizes its posterior probability $P(M^{1:2} \mid \Phi^{1:2}\ [C_m = L])$ its equivalent to minimizing the corresponding cost function under the perceptual constraints defined by GEPPETO. This complete the proof that the optimal control approach performed by GEPPETO is included as a special case of the Bayesian model.