$$\prod_{\ell=1}^{M}\Big[p(\mathbf{v}^{\ell,1},\ldots,\mathbf{v}^{\ell,n^\ell}\mid \boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_N,N)\Big]p(\boldsymbol{\mu}_1,\ldots\boldsymbol{\mu}_N\mid N)\,p(N) \tag{2}$$

$$=\prod_{\ell=1}^{M}\left[\sum_{\mathbf{a}^\ell\in A(n^\ell,N)}\left(\prod_{k=1}^{n^\ell}\underbrace{p(\mathbf{v}^{\ell,k}\mid \boldsymbol{\mu}_{a_k^\ell})}_{4}\right)\underbrace{p(\bar{V}(\mathbf{a}^\ell)\mid \boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_N,N)}_{3}\right]\underbrace{p(\boldsymbol{\mu}_1,\ldots\boldsymbol{\mu}_N\mid N)}_{2}\underbrace{p(N)}_{1}$$

Figure 2: Joint likelihood of M vowel systems under our deep generative probability model for continuous-space vowel inventories. Here language ℓ has an observed inventory of pronunciations $\{\mathbf{v}^{\ell,k} : 1 \leq k \leq n^\ell\}$, and $a_k^\ell \in [1, N]$ denotes a phone that might be responsible for the pronunciation $\mathbf{v}^{\ell,k}$. Thus, $\mathbf{a}^\ell$ denotes some way to jointly label all n^ℓ pronunciations with distinct phones. We must sum over all $\binom{N}{n^\ell}$ such labelings $\mathbf{a}^\ell \in A(n^\ell, N)$ since the true labeling is not observed. In other words, we sum over all ways $\mathbf{a}^\ell$ of completing the data for language ℓ. Within each summand, the product of factors 3 and 4 is the probability of the completed data, i.e., the joint probability of generating the inventory $\bar{V}(\mathbf{a}^\ell)$ of phones used in the labeling and their associated pronunciations. Factor 3 considers the prior probability of $\bar{V}(\mathbf{a}^\ell)$ under the DPP, and factor 4 is a likelihood term that considers the probability of the associated pronunciations.