INTRODUCTION

The human ability to perceive the direction of a sound source is partly the result of cues encoded in the sound reaching the eardrum after scattering off of the listener's anatomic features (torso, head, and outer ears). The frequency response of how sound is modified in phase and magnitude by such scattering is called the Head-Related Transfer Function (HRTF) [1] and is specific to each person. Knowledge of the HRTF allows for the reconstruction of realistic auditory scenes.

While the ability to collect HRTFs has existed for several years, and HRTFs of human subjects have been collected by different labs, there remain several issues with their widespread use. First, HRTFs show considerable variability between individuals. Second, each measurement facility seems to use an individual process to obtain the HRTF – using varying excitation signals, sampling frequencies, and more importantly measurement grids. The latter is a larger problem than may be initially thought, as the measurement grids are neither spatially uniform nor high resolution; time/cost issues and peculiarities of each measurement apparatus are limiting factors. Figure 1 illustrates a typical HRTF measurement grid). To overcome the grid problem, solutions via spherical interpolation techniques [2, 3, 4] are either performed on a per-frequency basis or in a principal component weight space over the measurement grid per subject. Personalization in a tensor-product principal component space was attempted by [5].

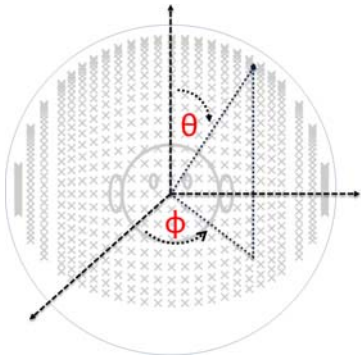


FIGURE 1: HRTF grid

A key development in statistical modeling has been the development of Bayesian methods, which learn from available data, and allow the incorporation of informative prior models. If HRTFs can be jointly modeled in their spatial-frequency domain under a Bayesian setting, then it might be possible to improve the ability to deal with these issues. Moreover, such a modeling can be done in an informative feature space, as is often done in speech-processing and image-processing. Spectral features (such as peaks and notches), used in [6, 7], are promising and correlate listening cues along specific directions (median plane) to anatomical features.

We propose a non-parametric spatial-frequency HRTF representation based on *sparse Gaussian process regression* (GPR) [8, 9] that addresses the aforementioned issues. Our model uses prior data (HRTF measurements) to infer HRTFs for previously unseen locations or frequencies for a single-subject. Our current work, not reported here, also seeks to extend this to modeling of inter-personal variability. The interpolation problem between the input spatial-frequency coordinate domain (ω, θ, ϕ) and the output HRTF measurement[1] $H(\omega, \theta, \phi)$ is non-parametric but does require the specification of a covariance model, which should reflect prior knowledge. Empirical observations [12, 13] suggest that the HRTF generally varies smoothly both over space and over frequency. In our model, the degree of smoothness is specified by the covariance model; this property also allows us to extract spectral features in a novel way via the derivatives of the interpolant. While our model can utilize the full collection of HRTFs belonging to the same subject for inference, it can also specify any subset of frequency-spatial inputs to jointly predict HRTFs at both original and new locations. Learning a subset of predictive HRTF directions as well as covariance function hyperparameters is an automatic process via marginal-likelihood optimization using Bayesian inference – a feature that other methods do not possess. HRTF data from the CIPIC database [13] are used in our interpolation, feature extraction, and importance sampling experiments.

[1] Through the current paper, we use the term *HRTF measurement* to refer exclusively to the magnitude part as HRTF can be reconstructed from magnitude response using min-phase transform and pure time delay [10, 11].

GP Regression

In a general regression problem, one predicts a scalar target variable y from a D-dimensional vector x of independent variables based on a collection of available observations (measurements). In the context of the HRTF interpolation problem, these variables respectively are the HRTF magnitude, and the angular and frequency coordinates. A common model assumes that the observations of y are generated by an unknown "latent" function $f(x)$, which is sought to be learnt probabilistically. The observations are assumed to be corrupted by additive (Gaussian) noise

$$y = f(x) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2), \tag{1}$$

where the noise term ϵ is zero centered with constant variance σ^2. Placing a GP prior distribution on $f(x)$ enables inference and encodes several useful properties such as local smoothness, stationarity, and periodicity. For any subset of inputs $X = [x_1, \ldots, x_N]$, the corresponding vector of function values $\mathbf{f} = [f(x_1), f(x_2), \ldots, f(x_N)]$ has a joint N-dimensional Gaussian distribution that is specified by the prior mean $m(x)$ and covariance $K(x_i, x_j)$ functions

$$f(x) \sim GP(m(x), K(x_i, x_j)), \quad m(x) = 0, \quad K(x_i, x_j) = \mathrm{Cov}(f(x_i), f(x_j)). \tag{2}$$

The joint distribution between N training and N_* test outputs y and f_* under the joint prior $p(f, f_*)$ is the $N + N_*$ dimensional normal distribution

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) + \sigma^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right), \quad \begin{matrix} K_{ff} = K(X,X), & \hat{K} = K_{ff} + \sigma^2 I, \\ K_{f*} = K(X, X_*), & K_{**} = K(X_*, X_*), \end{matrix} \tag{3}$$

where $K(X,X)$ and $K(X, X_*)$ are $N \times N$ and $N \times N_*$ sized covariance matrices with entries $K_{ij} = K(x_i, x_j)$ evaluated at all pairs of training and test inputs respectively. From Eq. 3 and marginalization over the function space $\mathbf{f}$, inference of any set of test outputs conditioned on the test inputs, training data, and training inputs yields the predictive distribution $P(f_*|X, y, X_*) \sim \mathcal{N}(\bar{f}_*, \mathrm{cov}(f_*))$ with the posterior mean and covariance functions

$$\bar{f}_* = E[f_*|X, y, X_*] = K_{f*}^T \hat{K}^{-1} y, \quad \mathrm{cov}(f_*) = K_{**} - K_{f*}^T \hat{K}^{-1} K_{f*}. \tag{4}$$

Thus, the interpolant $\bar{f}_*$ for inputs X_* in Eq. 4 is computed from the inversion of the covariance matrix $\hat{K}$ specified by the covariance function K, its hyperparameters, and control points (i.e. training outputs y). Model-selection is an $\mathrm{O}(N^3)$ runtime task of minimizing the gradient of the negative log-marginal likelihood function w.r.t. a hyperparameter Θ_i:

$$-\log p(y|X) = \frac{1}{2}\left(\log|\hat{K}| + y^T \hat{K}^{-1} y + N \log(2\pi)\right), \quad -\frac{\partial \log p(y|X)}{\partial \Theta_i} = \frac{1}{2}\left(\mathrm{tr}\left(\hat{K}^{-1} P\right) - y^T \hat{K}^{-1} P \hat{K}^{-1} y\right), \tag{5}$$

where $P = \partial \hat{K} / \partial \Theta_i$ is the matrix of partial derivatives.

Spatial-Frequency Grid GP Model

The GP prior covariance function encodes the assumed constraints on the latent function f. In particular, the degree of correlation between any subset of outputs (i.e., the smoothness, which is the desired interpolation property) is fully specified by the GP prior as a function over the input domain and a hyperparameter set. The frequency domain of a magnitude HRTF can be modeled by an *Ornstein-Uhlenbeck* (OU) [14] process, resulting in a GP with stationary auto-covariance $K_t(t_i, t_j) = \exp\left[-\frac{|t_i - t_j|}{\lambda}\right]$, and spectral density $K_\omega(\omega_i, \omega_j) = \frac{1}{\lambda^2 + (\omega_i - \omega_j)^2}$ functions where (t, ω) are time and frequency (inputs) and λ the hyperparameter. The λ term in the OU

process is the rate of mean reversion or drift to zero of a solution to a stochastic differential equation with standard Brownian motion. OU description agrees with the fact that the HRTF in time domain decays quickly to zero after the initial onset and is thus reasonable. For multiple HRTFs separated in the spherical domain, the joint magnitude responses for the same frequency can be modeled as a function of spherical coordinates (θ, ϕ) with the exponential function $K_s(\theta_i, \theta_j, \phi_i - \phi_j) = \exp\left[-\frac{Ch}{\ell^2}\right]$, with the "great circle" or chordal distance $Ch = 2\sqrt{\sin^2\left(\frac{\theta_j - \theta_i}{2}\right) + \sin\theta_i \sin\theta_j \sin^2\left(\frac{\phi_i - \phi_j}{2}\right)}$ on the unit sphere. In general, the *characteristic length-scale* hyperparameter ℓ is interpreted as the distance for function values to become uncorrelated w.r.t. an axis in the input domain.

We specify the GP covariance prior for the interpolant $\mathbf{f}$ as the *tensor product kernel* of exponential and OU spectral density functions with a global scaling hyperparameter α:

$$K(\theta_i, \theta_j, \phi_i - \phi_j, \omega_i - \omega_j) = \frac{\alpha^2}{\lambda^2 + (\omega_i - \omega_j)^2} \exp\left[-\frac{Ch_{ij}}{\ell^2}\right]. \tag{6}$$

For tractable inference and hyperparameter training, we decompose the Gram matrix K_{ff} into Kronecker tensor products (KTPs) $K_{ff} = K_s(X_s, X_s) \otimes K_\omega(X_\omega, X_\omega)$ for the multidimensional grid of inputs $x \in X$. Since HRTF collections contain magnitude responses for identical and uniform frequencies per spherical direction, the input domain can be defined as the Cartesian outer product $X = X_s \times X_\omega$ between direction and frequency sets $X_s \in \mathbb{R}^{m_s \times 2}$ and $X_\omega \in \mathbb{R}^{m_\omega \times 1}$ such that $N = m_s m_\omega$ is total number of input vectors. That is, the covariance matrices K_s and K_ω are independently computed from spatial and frequency inputs respectively; the inverse covariance matrix with additive white noise $\hat{K}$ can be efficiently evaluated as products of KTPs

$$\hat{K}^{-1} = U(Z + \sigma^2 I)^{-1} U^T, \quad U = U_s \otimes U_\omega, \quad Z = Z_s \otimes Z_\omega, \tag{7}$$

where matrices $K_s = U_s Z_s U_s^T \in \mathbb{R}^{m_s \times m_s}$ and $K_\omega = U_\omega Z_\omega U_\omega^T \in \mathbb{R}^{m_\omega \times m_\omega}$ have eigendecompositions. This "eigendecomposition trick" reduces inference and hyperparameter training time from $\mathrm{O}\left(N^3\right)$ to $\mathrm{O}\left((m_s^3 + m_\omega^3) + N(m_s + m_\omega)\right)$ and storage from $\mathrm{O}\left(N^2\right)$ to $\mathrm{O}\left(N + (m_s^2 + m_\omega^2)\right)$ as computing the Kronecker product expressions in Eqs. 4, 5 are efficient [15].

Spectral Extrema Extraction

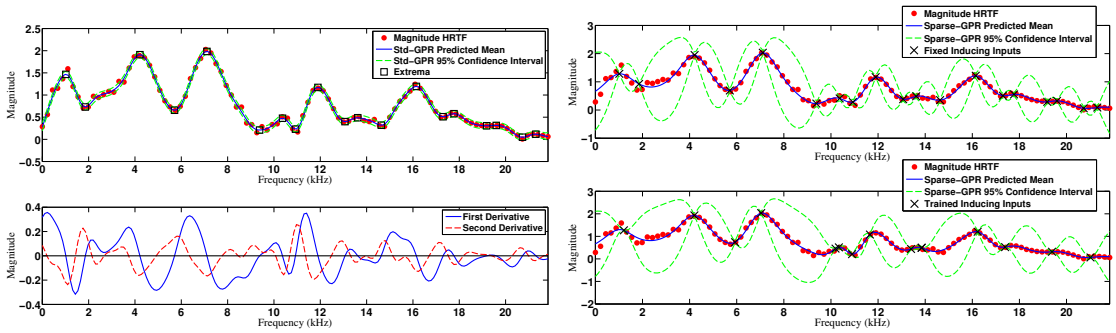
By definition, the spectral extrema correspond to the zero-crossing of the GP interpolant gradient. The surface gradient and the necessary covariance derivatives w.r.t. frequency ω_* are expressed in the form as

$$\begin{aligned} &\frac{\partial \bar{f}_*}{\partial \omega_*} = \left[\frac{\partial K_{1*}}{\partial \omega_*}, \ldots, \frac{\partial K_{N*}}{\partial \omega_*}\right] \hat{K}^{-1} y, \quad \frac{\partial^2 \bar{f}_*}{\partial \omega_*^2} = \left[\frac{\partial^2 K_{1*}}{\partial \omega_*^2}, \ldots, \frac{\partial^2 K_{N*}}{\partial \omega_*^2}\right] \hat{K}^{-1} y, \\ &\frac{\partial K_{i*}}{\partial \omega_*} = \frac{-2\alpha^2(\omega_* - \omega_i)}{(\lambda^2 + (\omega_* - \omega_i)^2)^2} e^{-Ch_{i*}/\ell^2}, \quad \frac{\partial^2 K_{i*}}{\partial \omega_*^2} = \frac{-2\alpha^2(\lambda^2 - 3(\omega_* - \omega_i)^2)}{(\lambda^2 + (\omega_* - \omega_i)^2)^3} e^{-Ch_{i*}/\ell^2}. \end{aligned} \tag{8}$$

Zero-crossings of the gradient signify spectral notches and peaks; these can be further distinguished by the sign of the second derivative. We use an iterative *Newton-Raphson* method:

$$\omega_{n+1} = \omega_n - \frac{\partial \bar{f}_{\omega_n}}{\partial \omega_n} \Big/ \frac{\partial^2 \bar{f}_{\omega_n}}{\partial \omega_n^2}, \quad |\omega_{n+1} - \omega_n| > \tau, \tag{9}$$

for locating zeros of the interpolant gradient in Eq. 8. The method was found to converge in a few iterations using a termination threshold of 10^{-5}. The initial guesses ω_0 are spaced uniformly in the frequency domain. Most prominent peaks and notches are extracted from the GP in Figure 2a after 50 iterations of hyperparameter training.



(A) Standard-GPR **(B)** Sparse-GPR

FIGURE 2: GPR and sparse-GPR of magnitude HRTF for subject 3, right ear, direction $(2.3562, 0)$. Spectral extrema are extracted from the zero-crossing of the predicted mean's gradient. The log-marginal likelihoods of standard-GP, sparse-GP with fixed inducing locations to spectral extrema, and trained inducing locations are $48.87, -18.18, -1.70$.

Sparse Grid GP Extension for Importance Sampling

To evaluate the predictive value of the spectral extrema to the original HRTF and to extract prominent directions from the spherical domain, we adopt sparse-GPR methods. A unified framework for sparse-GPR [9] is presented as a modification of the joint prior $p(f, f_*)$ that assumes conditional independence between function and predicted values f and f_* given a set of $M << N$ inducing inputs $\mathbf{u} = [u_1, \ldots, u_M]^T$ at inducing locations $X^{(u)}$ in the input domain. That is, the inducing pair $(X^{(u)}, \mathbf{u})$ represents a sparse set of latent inputs that can be optimized to infer the original data (X, y). One such sparse method is the deterministic training conditional (DTC) where the approximated joint prior $q(y, f_*) \sim p(y, f_*)$, after marginalizing out the inducing inputs $\mathbf{u}$, has the form

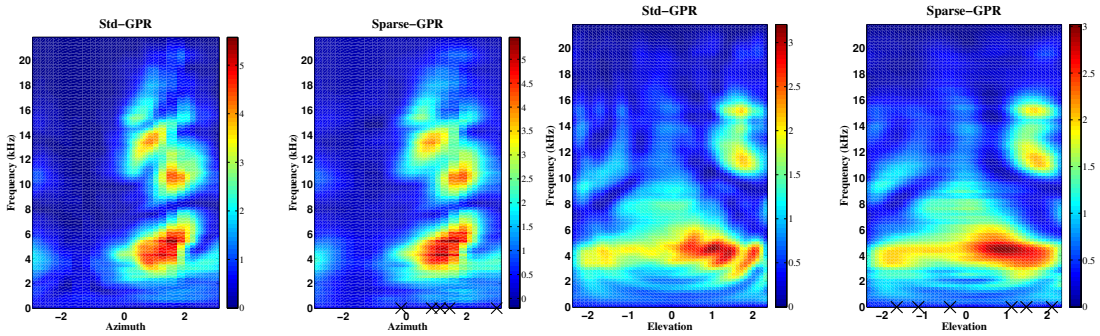
$$q(y, f_*) \sim \mathcal{N}\left(0, \begin{bmatrix} \hat{Q} & Q_{f*} \\ Q_{*f} & K_{**} \end{bmatrix}\right), \quad \hat{Q} = Q_{ff} + \sigma^2 I, \quad Q_{ab} = K_{au} K_{uu}^{-1} K_{ub}. \tag{10}$$

The low-rank matrix Q_{ff} in Eq. 10 is computed from $M \times M$ and $N \times M$ sized matrices $K_{uu} = K(X^{(u)}, X^{(u)})$ and $K_{fu} = K(X, X^{(u)})$ that approximates the original Gram matrix K_{ff}. For inference, the predictive distribution follows

$$\begin{aligned} q(f_*|y) &= \mathcal{N}(Q_{*f}(Q_{ff} + \sigma^2 I)^{-1} y, K_{**} - Q_{*f}(Q_{ff} + \sigma^2 I)^{-1} Q_{f*}) \\ &= \mathcal{N}(\sigma^{-2} K_{*u} \Sigma K_{uf} y, K_{**} - Q_{**} + K_{*u} \Sigma K_{u*}), \quad \Sigma = (\sigma^{-2} K_{uf} K_{fu} + K_{uu})^{-1}, \end{aligned} \tag{11}$$

which is handled in the covariance space spanned by the inducing locations $X^{(u)}$ as represented by matrix Σ. The sparse log-marginal likelihood function and its gradient w.r.t. hyperparameter Θ_i are analogous to Eq. 5 with the approximating matrix Q_{ff} replacing all instances of matrix K_{ff} and reexpressed in terms of matrix Σ (see [16] for the derivation). This allows hyperparameters and inducing locations $X^{(u)}$ (substituted as hyperparameters) to be trained via gradient descent of the objective negative sparse log-marginal likelihood function. Thus, the predictive value of any set of initial locations $X^{(u)}$ can be evaluated; training initial inducing locations set to spectral extrema frequencies (50 iterations) result in tighter prediction (see Figure 2b). In general, random initializations of the inducing locations converge to lower log-marginal likelihood minima than that of the spectral extrema.

For importance sampling, we extend the grid formulation to the sparse-GPR model by constraining the inducing locations to be a Cartesian outer product $X^{(u)} = X_s^{(u)} \times X_\omega^{(u)}$ between the original set of input frequencies X_ω and a sparse set of the directions $X_s^{(u)}$. By training only the inducing directions $X_s^{(u)}$, the method iteratively converges to a sparse set of spherical coordinates that reflect the modes of the original collection. We demonstrate this aspect along



(A) Horizontal Plane **(B)** Median Plane

FIGURE 3: GPR and sparse-GPR of HRTFs along horizontal and median planes for subject 3, right ear. Inducing inputs (X) representing prominent directions along the plane are learned in the sparse-GPR model after 50 iterations.

the wider collection of horizontal and median plane HRTFs in Figure 3. For tractable inference (inducing locations $X^{(u)}$ are sparse in only the spherical domain), a similar extension to Eq. 7 is made for matrix Σ. That is, the Kronecker structure for matrix Σ can be preserved via the eigendecomposition of KTP matrices $K_{uu} = UZU^T$ where $U = U_s \otimes U_\omega$ and $Z = Z_s \otimes Z_\omega$ along with a second set of eigendecompositions of KTP matrix $Z^{-1/2}U^T K_{uf} K_{fu} U Z^{-1/2} = \bar{U}\bar{Z}\bar{U}^T$. The matrix Σ can now be evaluated as KTPs

$$\Sigma = \sigma^2 \Omega(\bar{Z} + \sigma^2 I)^{-1}\Omega^T, \quad \Omega = UZ^{-1/2}\bar{U}, \quad \bar{U} = \bar{U}_s \otimes \bar{U}_\omega, \quad \bar{Z} = \bar{Z}_s \otimes \bar{Z}_\omega, \tag{12}$$

with reduced computational time and storage costs of $\mathrm{O}\left(m_s^{(u)^2}(m_s^{(u)^2} + m_s) + m_\omega^{(u)^2}(m_\omega^{(u)} + m_\omega)\right)$ and $\mathrm{O}\left(m_s^{(u)}(m_s^{(u)} + m_s) + m_\omega^{(u)}(m_\omega^{(u)} + m_\omega)\right)$ respectively.

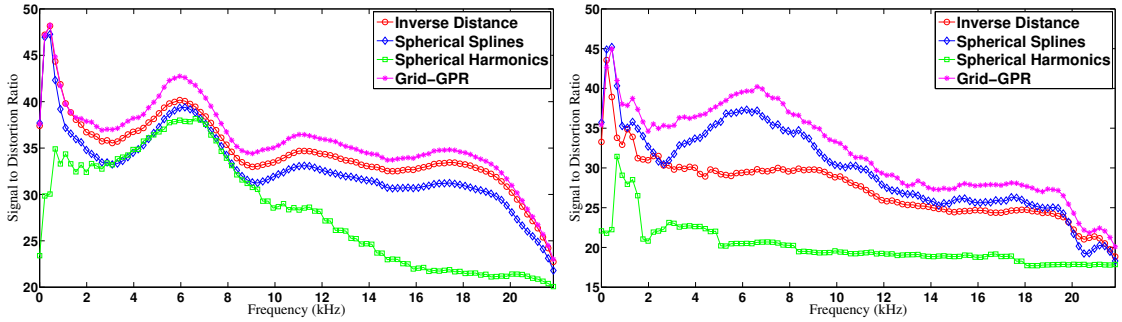
INTERPOLATION AND EXTRAPOLATION

From the CIPIC database, HRTFs from the 45 subjects are computed using DFT of the 200-sample-long impulse response and the magnitude of the first 100 bins are taken due to symmetry. The measure of the interpolation error is the signal-to-distortion ratio (SDR): $SDR_\omega = 10\log_{10}\frac{\sum_{i=1}^{N_*} H_\omega(\theta_i,\phi_i)^2}{\sum_{i=1}^{N_*}(H_\omega(\theta_i,\phi_i) - \hat{H}_\omega(\theta_i,\phi_i))^2}$, where H and $\hat{H}$ are the true and predicted magnitude responses at frequency ω over N_* prediction directions respectively. GP hyperparameters are trained after 50 iterations of gradient descent.

In the first experiment, randomly-chosen half of 1250 CIPIC grid directions comprise the "measurement" data set; the HRTF interpolation is done to obtain the predicted data for the rest of directions and the prediction error is computed. The task measures how well the global interpolant is suited for the case of arbitrary measurement grid. SDR plots for each interpolation methods are shown in Figure 4a; high SDR corresponds to a good reconstruction. For the spherical spline method [17], the default parameters for smoothing and expansion terms are used. For the spherical harmonic fitting [18], truncated SVD regularization method is used as described. The results show that our GP model with Bayesian model selection uniformly outperforms all other methods in the $2-20$ kHz frequency range by several SDRs.

For the second experiment, we simulate missing data in a large spatial area (an open hole task [19]). We remove all measurements that lie above certain horizontal plane (specifically having spherical incident angle $\theta < \pi/5$), essentially cutting off the top portion of the sphere of direction. There are a total of 147 measurement directions in the cut-off area. The interpolant is computed over the rest of measurements, is evaluated in the hole area, and the SDR is computed. This task simulates the HRTF prediction in the areas where data is not available

(such as the bottom hole in most HRTF measurement grids). The SDR across each frequency are shown in Figure 4b. Our GPR model exhibits the greatest gains along the $2-10$ kHz frequency range and consistently outperforms the global spline method in other frequencies.

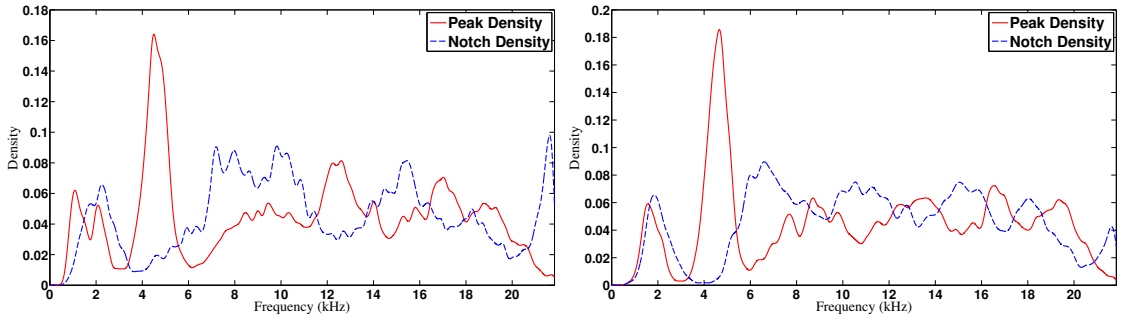


(A) Interpolation Accuracy **(B)** Extrapolation Accuracy

FIGURE 4: Cumulative SDRs (dB) for the interpolation (random partition) and extrapolation (missing hole) tasks across all subject's right ear HRTFs. Larger SDRs per frequency indicate smaller error.

ANALYSIS OF SPECTRAL EXTREMA AND DIRECTIONAL SAMPLING

Having available an efficient algorithm for HRTF feature extraction, we now seek to study the distribution of these features across subjects, over the space of directions. Two collections of spectral extrema and sampling directions are extracted from HRTFs belonging to horizontal and median planes across all subject, right ears via the sparse-GPR model. For the collection of spectral extrema, two multi-modal kernel density estimations (KDEs) with Gaussian kernels and optimized bandwidth [20] are fitted over the sets of peak and notch frequencies in Figure 5. The horizontal plane extrema indicate 2 modes of notches beyond the 4 kHz range along frequencies $7-11$ and $14-16$ kHz that are not well-reflected in the corresponding peak densities. The median plane extrema indicate 4 modes of notches beyond the 4 kHz range centered along frequencies $6, 11, 15, 18$ kHz that have analogous densities in the peaks shifted by $+2$ kHz. Both models share similar distributions in the lower frequency ranges attributed to the initial torso and head reflections.

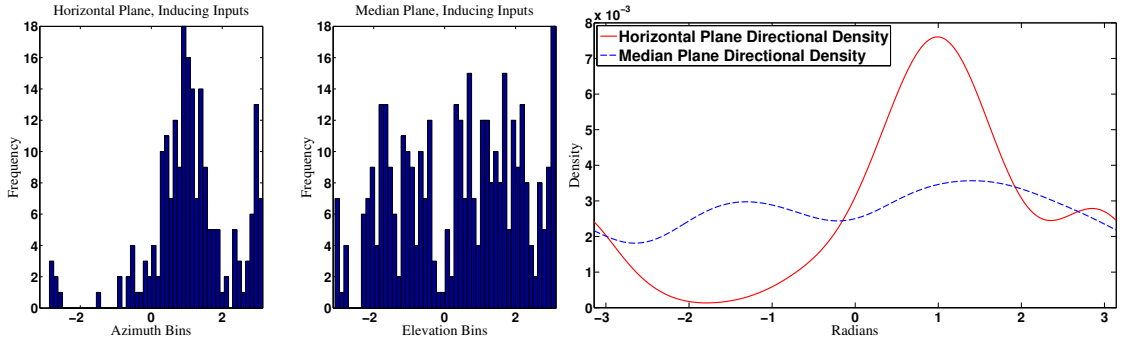


(A) KDE of Horizontal Plane Extrema **(B)** KDE of Median Plane Extrema

FIGURE 5: Kernel density estimation (Gaussian) of pooled spectral extrema for GPs trained on horizontal and median plane HRTFs across all subjects, right ears.

For the collection of prominent sampling directions, a circular-KDE with optimized Gaussian kernel bandwidth [21] are fitted over azimuth and elevation radians in the horizontal and median planes. Two bimodal distributions are visible in Figure 6 for both planes. In the

horizontal plane, the two modes are distinguished by directions facing the pinna and behind the right-ear-flap but not obscured by the head shadow; pinna features are exposed along the dominant 1 radian mode mean. For the median plane, the two modes have near-orthogonal means along $-\pi/2, \pi/2$ radians and represent elevation cues in front and behind the head. The similar densities suggest a comparable number of relevant features that may account for difficulties in front-back discernment in sound localization.



(A) Histogram of Inducing Directions **(B)** Circular-KDE of Inducing Directions

FIGURE 6: Kernel density estimation (Gaussian) of pooled inducing directions trained from sparse-GPR models on horizontal and median plane HRTFs across all subjects, right ears. Inducing inputs represent predictive sampling directions to the original HRTF collection.

CONCLUSIONS

We have presented a joint spatial-frequency sparse GPR model with separable covariances for single-subject HRTF representation and interpolation. The proposed method achieved better interpolation/extrapolation accuracy in comparison with other existing spherical interpolation methods and provided a particularly efficient method for the extraction of HRTF spectral extrema and prominent sampling directions. This allowed us to study the HRTF in feature space efficiently. A first step in this feature space study was the modeling of the variation of HRTFs across multiple subjects; density estimations of these features indicate clustering behavior in both domains along the horizontal and median planes. Future work will investigate the effects of manipulating this feature space in listening tests and their relation to anthropometry in the personalization problem.

REFERENCES

[1] J. Blauert, *Spatial hearing: the psychophysics of human sound localization* (MIT Press, Cambridge, Massachusettes) (1997).

[2] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of HRTFs", in *IEEE ICASSP*, volume 4, 45–48 (Montreal, QC, Canada) (2004).

[3] S. M. Robeson, "Spherical methods for spatial interpolation: Review and evaluation", Cartography and Geographic Information Science **24**, 3–20 (1997).

[4] E. M. Wenzel and S. H. Foster, "Perceptual consequences of interpolating head-related transfer functions during spatial synthesis", in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (1993).

[5] G. Grindlay and M. Vasilescu, “A multilinear (tensor) framework for HRTF analysis and synthesis”, in *IEEE ICASSP* (2007).

[6] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, “Extracting the frequencies of the pinna spectral notches in measured head related impulse responses”, Journal of Acoustical Society of America **118**, 364–374 (2005).

[7] V. R. Algazi, C. Avendano, and R. O. Duda, “Elevation localization and head-related transfer function analysis at low frequencies”, Journal of the Acoustical Society of America **109**, 1110–1122 (2001).

[8] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, Massachusettes) (2006).

[9] J. Quinonero-Candela and C. E. Rasmussen, “A unifying view of sparse approximate Gaussian process regression”, Journal of Machine Learning Research **6**, 1939–1959 (2005).

[10] D. J. Kistler and F. L. Wightman, “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction”, Journal of Acoustical Society of America **91**, 1637–1647 (1992).

[11] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, “Sensitivity of human subjects to head-related transfer-function phase spectra”, Journal of the Acoustical Society of America **105**, 2821–2840 (1999).

[12] A. Kulkarni and H. S. Colburn, “Role of spectral detail in sound-source localization”, Nature **396**, 747–749 (1998).

[13] V. R. Algazi, R. O. Duda, and C. Avendano, “The CIPIC HRTF Database”, in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 99–102 (New Paltz, NY) (2001).

[14] G. E. Uhlenbeck and L. S. Ornstein, “On the theory of Brownian motion”, Phys. Rev **36**, 823–841 (1930).

[15] Y. Saatci, “Scalable inference for structured Gaussian process models”, Ph.D. thesis, University of Cambridge (2011).

[16] J. Quinonero-Candela, “Learning with uncertainty - Gaussian processes and relevance vector machines”, Ph.D. thesis, Technical University of Denmark (2004).

[17] J. Kayser and C. E. Tenke, “Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks.”, Clinical Neurophysiology **117**, 348–368 (2006).

[18] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, “Regularized HRTF fitting using spherical harmonics”, in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 257–260 (2009).

[19] W. Zhang, R. A. Kennedy, and T. D. Abhayapala, “Iterative extrapolation algorithm for data reconstruction over sphere”, in *IEEE ICASSP*, 3733–3736 (2008).

[20] Z. Botev, J. Grotowski, and D. Kroese, “Kernel density estimation via diffusion”, Annals of Statistics **38**, 2916–2957 (2010).

[21] B. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall/CRC, London) (1998).