

GAUSSIAN PROCESS DATA FUSION FOR HETEROGENEOUS HRTF DATASETS

Yuancheng Luo¹ Dmitry N. Zotkin² Ramani Duraiswami¹ *

¹ University of Maryland, Department of Computer Science, yluo1@umd.edu, ramani@umiacs.umd.edu

² University of Maryland, Institute for Advanced Computer Studies, dz@umiacs.umd.edu

ABSTRACT

Head-Related Transfer Function (HRTF) measurement and extraction are important tasks for personalized-spatial audio. Many laboratories have their own apparatuses for data-collection but few studies have compared their results to a common subject or have modeled inter-dataset variances. We present a Bayesian fusion method based on Gaussian process (GP) modeling of joint *spatial-frequency* HRTFs over different spherical-measurement grids. Neumann KU-100 dummy HRTFs from 7 labs in the “Club Fritz” study are compared and *fused* to each other based on learning a set of transformations from the GP data-likelihood and covariance assumptions; parameter and hyperparameter training is automatic. Experimental results show that fused models for horizontal and median-plane HRTFs generalize the datasets better than pre-transformed ones.

Index Terms— Gaussian Process Regression, Data Fusion, Kronecker Product, Equalization, Windowing

1. INTRODUCTION

Human sound localization in 3D is based partially on the monaural and binaural cues that arise from a source’s acoustic wave scattering off of the listener’s anatomy (torso, head, and outer ears). The ratio of the Fourier Transform of this wave measured at the listener’s eardrum to that which would have been present at the head-center location in the absence of the listener is called the Head-Related Transfer Function (HRTF) [1] and is unique to each person due to differences in anthropometry. Knowing a person’s HRTF enables one to synthesize life-like auditory scenes.

While many research labs have measured HRTFs for human listeners, very few comparisons have been made between common subjects; how inter-lab HRTF datasets can be compared and then combined for greater predictive performance remain open questions. Indeed, several sub-problems from the selection of a common subject, the correspondence between different HRTF measurement grids, and the class of transformations that models inter-lab HRTF variances must be answered. Causes of inter-lab variances may arise from measurement noise, positioning errors, non-omnidirectional directivity patterns, temperature dependent equipment transfer function drifts, from incompatible free-field equalizations etc. Learning a set of transformations that captures the inter-lab dataset variances for a common subject may generalize the variances between measurement processes, which can then be applied to HRTFs of subjects unique to each database. This enables developments in the study of HRTF statistics in a Bayesian setting where larger populations provide greater predictive value.

To answer such questions, a large round-robin activity was organized [2]. We made use of this rather unique “Club Fritz” database of Neumann KU-100 dummy HRTF measurements from 7

labs. Each dataset is defined over separate measurement grids in the spherical coordinate domains shown in Fig. 1 with corresponding HRTF measurements¹ in those directions. The measurement grids vary substantially over the sphere, which makes difficult to match measurements between datasets.

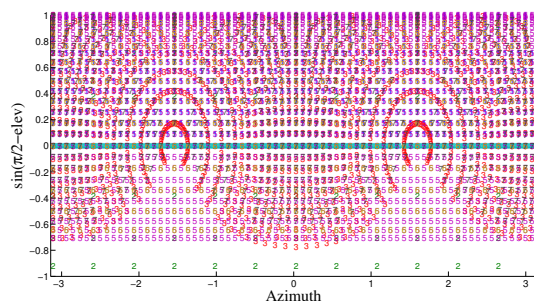


Figure 1: Mercator projection (spherical domain $(\theta, \phi) \rightarrow$ cylinder) of measurement grids for “Club Fritz” Neumann HRTFs. For anonymity, source institutions are indicated by lab numbers.

We propose a joint spatial-frequency Bayesian model where same-subject HRTF measurements $H_i(\theta_j, \phi_j, \omega_j)$ for all datasets \mathcal{D}_i are assumed to be drawn from the same latent function f (non-parametric) with GP priors; GPs are random fields where any finite subset of random variables are jointly Gaussian and defined by prior mean and covariance functions [3]. The covariance assumption asserts that topologically similar direction-frequency inputs produce similar realizations of magnitude HRTF-frequency responses with high probability; this coincides with the observation that magnitude HRTFs are often smooth in the spatial-frequency domains [4]. See sections 2 and 3 for a review of GPs and our covariance model.

While the smoothness assumption is consistent for same-subject HRTFs collected within a lab [4], this may not be true for inter-lab datasets as shown by HRTF measurements in Fig. 3; HRTF measurements from different labs are inconsistent with a quantitative data-likelihood measure given the GP priors as defined in section 4. We address the issue of recovering the source data discrepancies by learning two types of transformations that improves the combined data-likelihood; the impulse response spectrum and time-domains are point-wise multiplied by unknown filters to be learned (equalization and windowing) in sections 4.1 and 4.2 respectively. One dataset in every combined model remains constant; this stops

¹ HRTF measurements are preprocessed by recovering their minimum-phase Head-Related Impulse Responses (HRIR) to remove time-delay, resampling the HRIR to 44100 kHz, taking the magnitude of the discrete Fourier transform of the first 256 taps, truncating to the 0 – 18 kHz range, and scaling the magnitude range to (0, 1). We use the term *HRTF measurement* to refer exclusively to HRTF magnitude.

*Partial support of NSF award IIS-1117716 is gratefully acknowledged.

any transformation from degenerating towards the prior mean and interpreted as *fusing* source datasets $\mathcal{D}_{i \neq t}$ to a fixed-target dataset \mathcal{D}_t . Experiments show that transformed datasets generalize target datasets better than pre-transformed cases in section 5.

2. GAUSSIAN PROCESS 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 a parametric model, the problem is one of estimating the parameters of the model based on the data. When a parametric model is unknown, a common Bayesian approach of inference assumes that the observation y is generated by an unknown latent function $f(x)$ and is corrupted by additive (Gaussian) noise $y = f(x) + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma^2)$, where the noise term ϵ is zero centered with constant variance σ^2 . Placing a GP prior distribution on the latent function $f(x)$ enables inference and encodes several useful properties such as local smoothness, stationarity, and periodicity. For any subset of inputs $X = [x_1, \dots, x_N]$, the corresponding vector of function values $\mathbf{f} = [f(x_1), f(x_2), \dots, f(x_N)]$ has a joint N -dimensional Gaussian distribution that is specified by the prior mean $m(x)$ and covariance $K(x_i, x_j)$ functions given by

$$f(x) \sim \mathbf{GP}(m(x), K(x_i, x_j)), \quad m(x) = 0, \quad (1)$$

$$K(x_i, x_j) = \mathbf{cov}(f(x_i), f(x_j)).$$

For N training outputs y and N_* test outputs f_* , we define the Gram matrix $\hat{K} = K_{ff} + \sigma^2 I$ as the pair-wise covariance evaluations between training and test predictors given by matrices $K_{ff} = K(X, X) \in \mathbb{R}^{N \times N}$, $K_{f*} = K(X, X_*) \in \mathbb{R}^{N \times N_*}$, and $K_{**} = K(X_*, X_*) \in \mathbb{R}^{N_* \times N_*}$. The joint-prior is given by

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X, X) + \sigma^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right). \quad (2)$$

GP inference is a marginalization over the function space \mathbf{f} , which expresses the set of test outputs conditioned on the test inputs, training data, and training inputs as a normal distribution $P(f_* | X, y, X_*) \sim \mathcal{N}(\bar{f}_*, \mathbf{cov}(f_*))$ given by

$$\bar{f}_* = E[f_* | X, y, X_*] = K_{f*}^T \hat{K}^{-1} y, \quad (3)$$

$$\mathbf{cov}(f_*) = K_{**} - K_{f*}^T \hat{K}^{-1} K_{f*}.$$

For inference, we treat the posterior mean \bar{f}_* for inputs X_* in Eq. 3 as the interpolant, which is parameterized by the covariance function K , its hyperparameters, and the supports (X, y) .

3. PRIOR COVARIANCE AND MEAN MODELS

The GP prior covariance function establishes the degree of correlation between any subset of realizations (magnitude frequency responses) $f(\theta, \phi, \omega)$ as a function of input spherical coordinates and frequencies. The product of separable covariance functions is investigated for HRTF measurements [4], which takes advantage of the HRTF parameterization by Cartesian outer-product inputs $X = X^{(\theta\phi)} \times X^{(\omega)}$; the Gram matrix K_{ff} decomposes into Kronecker tensor products (KTP) $K_{ff} = K_1(X^{(\theta\phi)}, X^{(\theta\phi)}) \otimes K_2(X^{(\omega)}, X^{(\omega)})$ of separable covariance evaluations in the spatial $X^{(\theta\phi)}$ and spherical $X^{(\omega)}$ domains respectively. Efficient KTP computations for GP inference and hyperparameter training can also be found in [5].

We adopt the stationary covariance function $K(C_h, r) = K_1(C_h)K_2(r)$ of the product of the Matérn ($\nu = 3/2$) covariance function [3] for Chordal distance $C_h = 2\sqrt{\sin^2\left(\frac{\theta_j - \theta_i}{2}\right) + \sin\theta_i \sin\theta_j \sin^2\left(\frac{\phi_i - \phi_j}{2}\right)}$ and the spectral density of the *Ornstein-Uhlenbeck* (OU) auto-covariance [6] for frequency distance $r = |\omega_i - \omega_j|$ given by

$$K_1(C_h) = \left(1 + \frac{\sqrt{3}C_h}{\ell}\right) e^{-\frac{\sqrt{3}C_h}{\ell}}, \quad K_2(r) = \frac{2\alpha^2 \lambda^2}{\lambda^2 + r^2}. \quad (4)$$

Hyperparameters α , λ , and ℓ are the global-scale factor, the rate of mean drift to 0 in the OU process, and the characteristic *length-scale* between zero-crossings of 1D functions drawn from the GP prior with mean 0 respectively. Other combinations of covariance products including Matérn $\nu = \{1/2, 5/2, \infty\}$ lead to lower data-likelihood estimates in Eq. 5 for individual datasets \mathcal{D}_i after hyperparameter training. The prior mean $m(x) = 0$ ensures that the posterior mean reverts to 0 at the frequency limits; the interpolant \bar{f} representing the transfer function has finite integration.

4. DATA FUSION

We first establish notation as follows: For T HRTF datasets, let inputs $X = \{X^{\{1\}}, \dots, X^{\{T\}}\}$ correspond to observations $y = [y^{\{1\}}; \dots; y^{\{T\}}]$ by the dataset pairings $\mathcal{D}_i = \{X^{\{i\}}, y^{\{i\}}\}$; datasets have different measurement grids but share the same frequencies, which serve as valid inputs for the covariance model in Eq. 4. Indeed, a single latent function f can be specified over the full dataset (X, y) but will not be consistent with GP priors; large variations between datasets, even after free-field equalization, motivate a second data-transformation step proposed between a set of source datasets $\mathcal{D}_{i \neq t}$ and a fixed-target dataset \mathcal{D}_t . The setup encourages transformations to co-vary source datasets with the target dataset without specifying priors over the transformation parameters; the former uses target datasets as priors, the latter would require knowledge of inter-lab variance which we aim to learn.

Formally, let function $g_t(y)$ with parameters $\Theta^{\{t\}}$ transform the observation-vector y for fixed-observations $y^{\{t\}}$ and input-vector X . If GP prior mean and covariance functions are specified over a latent function f_t with isotropic noise over transformed observations $g_t(y)$, then the data-likelihood of $g_t(y)$ is the probability of having been drawn from the modified joint-prior normal distribution in Eq. 2. The related negative log-marginal likelihood objective function and its partial derivatives w.r.t. covariance hyperparameter $\Theta_i^{\{K, t\}}$ and transform-parameters $\Theta_i^{\{t\}}$ are given by

$$-\mathcal{L}_t = \frac{1}{2} \left(\log |\hat{K}| + g_t(y)^T \gamma + N \log(2\pi) \right),$$

$$-\frac{\partial \mathcal{L}_t}{\partial \Theta_i^{\{K, t\}}} = \frac{1}{2} \left(\mathbf{tr} \left(\hat{K}^{-1} \frac{\partial \hat{K}}{\partial \Theta_i^{\{K, t\}}} \right) - \gamma^T \frac{\partial \hat{K}}{\partial \Theta_i^{\{K, t\}}} \gamma \right), \quad (5)$$

$$-\frac{\partial \mathcal{L}_t}{\partial \Theta_i^{\{t\}}} = \gamma^T \frac{\partial g_t(y)}{\partial \Theta_i^{\{t\}}}, \quad \gamma = \hat{K}^{-1} g_t(y).$$

The closed-form derivatives provide automatic model-selection and transform-parameter learning by gradient descent methods. Several transform-functions g_t with physical interpretations are considered.

4.1. Equalization-Transform

The *equalization-transform* applies a common separable filter (diagonal-matrix vector product) to measurements in the direction-

frequency domains given by

$$g_t(y) = \mathbf{diag} \left[\Phi_t^{\{1\}}, \dots, \Phi_t^{\{t-1\}}, 1^{N_t}, \Phi_t^{\{t+1\}}, \dots, \Phi_t^{\{T\}} \right] y, \\ \Phi_t^{\{i\}} = \left(\Theta^{\{t,i,1\}} \otimes \Theta^{\{t,i,2\}} \right) \in \mathbb{R}^{1 \times |X_{\theta\phi}^{\{i\}}| \times |X_{\omega}^{\{i\}}|}, \quad (6)$$

where $1^{N_t} \in \mathbb{R}^{N_t \times 1}$ is the vector of ones that preserve dataset $Y^{\{t\}}$ via identity transformation. Task-independent transformations $\Phi_t^{\{i\}}$ are Kronecker diagonal-products between filter weights (parameters) $\Theta^{\{t,i,1\}}$ and $\Theta^{\{t,i,2\}}$ along the direction-frequency domains. For notation, we denote the partial derivatives of $g_t(y)$ w.r.t. parameters as $u = \partial g_t(y) / \partial \Theta_j^{\{t,i,1\}}$ and $v = \partial g_t(y) / \partial \Theta_j^{\{t,i,2\}}$ given by

$$u = \mathbf{diag} \left[0^{N_1}, \dots, 0^{N_{t-1}}, \frac{\partial \Phi_t^{\{i\}}}{\partial \Theta_j^{\{t,i,1\}}}, 0^{N_{t+1}}, \dots, 0^{N_T} \right] y, \\ v = \mathbf{diag} \left[0^{N_1}, \dots, 0^{N_{t-1}}, \frac{\partial \Phi_t^{\{i\}}}{\partial \Theta_j^{\{t,i,2\}}}, 0^{N_{t+1}}, \dots, 0^{N_T} \right] y, \quad (7) \\ \frac{\partial \Phi_t^{\{i\}}}{\partial \Theta_j^{\{t,i,1\}}} = e_j \otimes \Theta^{\{t,i,2\}}, \quad \frac{\partial \Phi_t^{\{i\}}}{\partial \Theta_j^{\{t,i,2\}}} = \Theta^{\{t,i,1\}} \otimes e_j,$$

where e_i the i^{th} column of the identity matrix. If $\Theta^{\{t,i,1\}} = 1^{|X_{\theta\phi}^{\{i\}}|}$ is fixed, then optimizing for $\Theta^{\{t,i,2\}}$ is interpreted as equalizing all magnitude HRTFs in dataset \mathcal{D}_i by a common filter. Conversely, fixing $\Theta^{\{t,i,2\}} = 1^{|X_{\omega}^{\{i\}}|}$ and optimizing for $\Theta^{\{t,i,1\}}$ uniquely scales the full magnitude spectrum for each input direction in dataset \mathcal{D}_i . Since the objective function \mathcal{L}_t is quadratic w.r.t. each $\Theta_j^{\{t,i,1\}}$ and $\Theta_j^{\{t,i,2\}}$, their minimums are given when

$$\Theta_j^{\{t,i,1\}} = -\frac{g_t(y)_u^T \hat{K}^{-1} u}{u^T \hat{K}^{-1} u}, \quad \Theta_j^{\{t,i,2\}} = -\frac{g_t(y)_v^T \hat{K}^{-1} v}{v^T \hat{K}^{-1} v}, \quad (8)$$

where $g_t(y)_u = g_t(y) - \Theta_j^{\{t,i,1\}} u$ and $g_t(y)_v = g_t(y) - \Theta_j^{\{t,i,2\}} v$ zero-out the entry with the parameter as a leading coefficient. Optimizing parameters thus monotonically increase the log-marginal likelihood objective function \mathcal{L}_t with fast convergence.

4.2. Window-Transform

The *window-transform* simulates windowing in the time-domain via a symmetric Toeplitz-matrix vector product in the direction-frequency domain given by

$$g_t(y) = \mathbf{bdg} \left[\Phi_t^{\{1\}}, \dots, \Phi_t^{\{t-1\}}, I_{N_t}, \Phi_t^{\{t+1\}}, \dots, \Phi_t^{\{T\}} \right] y, \quad (9) \\ \Phi_t^{\{i\}} = \mathbf{Tp} \left(\Theta^{\{t,i,1\}} \right) \otimes \mathbf{Tp} \left(\Theta^{\{t,i,2\}} \right),$$

where $\mathbf{bdg} [A_1, A_2]$ generates a block-diagonal matrix with diagonal elements as square matrices A_1, A_2 and 0's off-diagonal. Task-independent transformations $\Phi_t^{\{i\}}$ are Kronecker products of symmetric-Toeplitz matrices $\mathbf{Tp}(a)_{jk} = a_{|j-k|+1}$ generated from weights (parameters) $\Theta^{\{t,i,1\}}$, and $\Theta^{\{t,i,2\}}$ of the same size as Eq. 6. Optimizing parameters w.r.t. the objective function \mathcal{L}_t can be interpreted as learning a set of discrete and symmetric point-spread functions from sources to target datasets. The partial derivatives

$u = \partial g_t(y) / \partial \Theta_j^{\{t,i,1\}}$ and $v = \partial g_t(y) / \partial \Theta_j^{\{t,i,2\}}$ are given by

$$u = \mathbf{bdg} \left[0_{N_1}, \dots, 0_{N_{t-1}}, \frac{\partial \Phi_t^{\{i\}}}{\partial \Theta_j^{\{t,i,1\}}}, 0_{N_{t+1}}, \dots, 0_{N_T} \right] y, \\ v = \mathbf{bdg} \left[0_{N_1}, \dots, 0_{N_{t-1}}, \frac{\partial \Phi_t^{\{i\}}}{\partial \Theta_j^{\{t,i,2\}}}, 0_{N_{t+1}}, \dots, 0_{N_T} \right] y, \quad (10)$$

where $0_{N_i} \in \mathbb{R}^{N_i \times N_i}$ is the zero-matrix, $\partial \Phi_t^{\{i\}} / \partial \Theta_j^{\{t,i,1\}} = \mathbf{Tp}(e_j) \otimes \mathbf{Tp}(\Theta^{\{t,i,2\}})$ and $\partial \Phi_t^{\{i\}} / \partial \Theta_j^{\{t,i,2\}} = \mathbf{Tp}(\Theta^{\{t,i,1\}}) \otimes \mathbf{Tp}(e_j)$. The local minimum has the closed-form expression identical to Eq. 8 after substitutions, which allows multiple parameters to quickly converge during joint-optimization.

The composition of the window-transform $g_{w,t}$ and the equalization-transform $g_{e,t}$ given by $g_{w,t}(g_{e,t}(y))$ can easily be optimized by modifying Eq. 8: For parameters $\Theta_w^{\{t\}} \in g_{w,t}$, observation vector y is replaced with $g_{e,t}(y)$ in Eqs. 9 and 10. For parameters $\Theta_e^{\{t\}} \in g_{e,t}$, both the partial derivatives $\partial \Phi_t^{\{i\}} / \partial \Theta_j^{\{t,i,1\}}$ and $\partial \Phi_t^{\{i\}} / \partial \Theta_j^{\{t,i,2\}}$ in Eq. 7 are left-multiplied by parameters $\Phi_t^{\{i\}}$ from Eq. 9. The front-to-end process is summarized in Fig. 2.

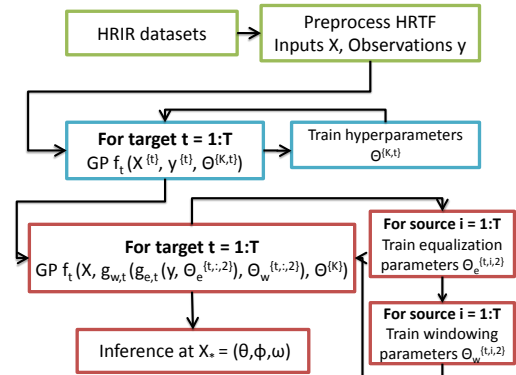


Figure 2: Flow-diagram for GP model-selection and transform-parameter learning. Free-field equalized HRIRs along the horizontal and median-planes are preprocessed¹ into HRTF measurements. Covariance hyperparameters $\Theta^{\{K,t\}}$ for interpolant f_t are independently trained for 100 iterations via Eq. 5 for GPs specified over the target dataset \mathcal{D}_t only. Interpolant f_t specified over transformed datasets $g_{w,i}(g_{e,i}(y))$ with trained GP prior covariance hyperparameters $\Theta^{\{K,t\}}$ have initial transform-parameters $\Theta_e^{\{t,i,1\}} = 1^{|X_{\theta\phi}^{\{i\}}|}$, $\Theta_e^{\{t,i,2\}} = 1^{|X_{\omega}^{\{i\}}|}$, $\Theta_w^{\{t,i,1\}} = e_1$, and $\Theta_w^{\{t,i,2\}} = e_1$. Frequency transform-parameters $\Theta_e^{\{t,i,2\}}$ and $\Theta_w^{\{t,i,2\}}$ are trained for 5 iterations via Eq. 8 over all source datasets \mathcal{D}_i .

5. EXPERIMENTS

The generalization error of each fused GP interpolant \bar{f}_t is the prediction error at the original inputs $X^{\{t\}}$ given by the signal-to-distortion ratio (SDR) over spatial-frequency inputs $(\theta_i, \phi_i, \omega)$:

$$\text{SDR}_{\omega} = 10 \log_{10} \frac{\sum_{i=1}^{N^*} H_{\omega}(\theta_i, \phi_i)^2}{\sum_{i=1}^{N^*} (H_{\omega}(\theta_i, \phi_i) - \hat{H}_{\omega}(\theta_i, \phi_i))^2}, \quad (11)$$

where $H_{\omega}(\theta_i, \phi_i) = y_{\omega, \theta_i, \phi_i}$ is the magnitude HRTF observation and $\hat{H}_{\omega}(\theta_i, \phi_i) = \bar{f}_{\omega, \theta_i, \phi_i}$ is the posterior mean in Eq. 3. Indeed, the SDRs for interpolant \bar{f}_t specified over the transformed

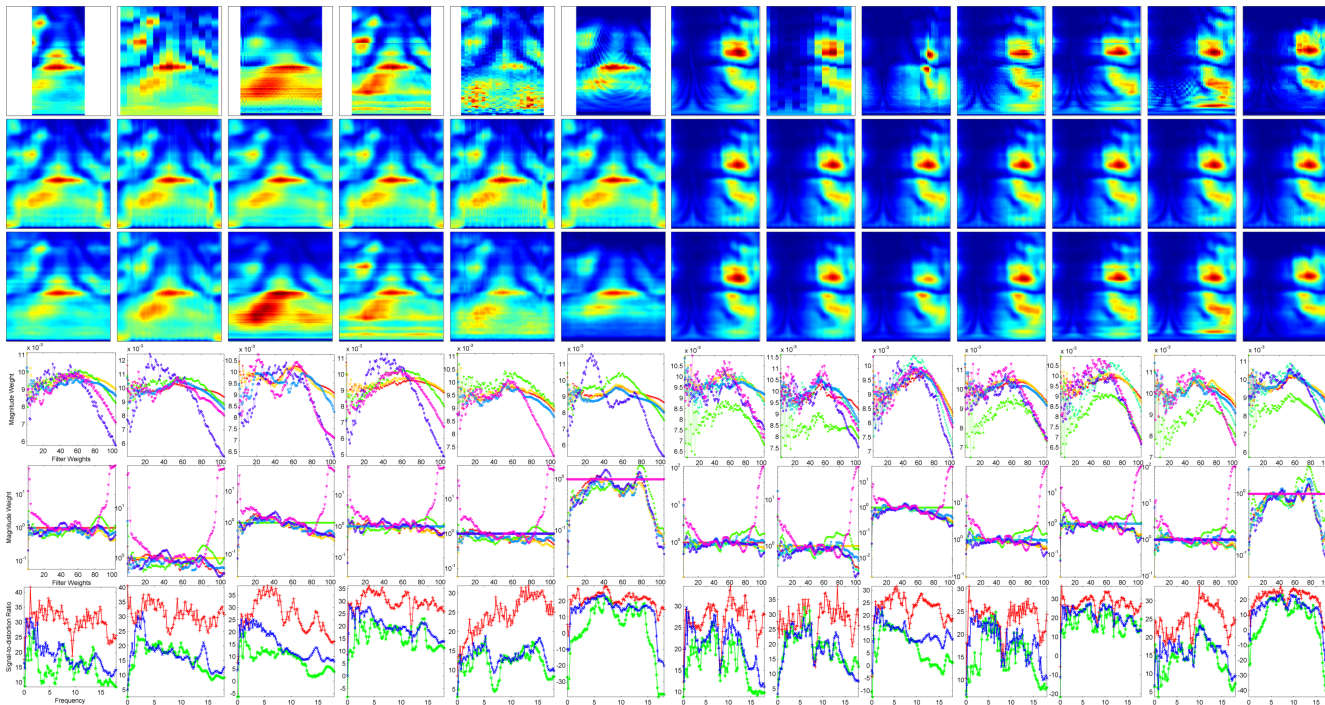


Figure 3: The “busy” plot shows all lab results combined and best viewed zoomed online. Row 1: Original club Fritz right-ear HRTF measurements for labs 1, 2, 3, 5, 6, 7 on meridional plane and labs 1 – 7 on horizontal plane (x-axis $-\pi < \theta, \phi < \pi$ and y-axis $0 < \omega < 18$ kHz). Plots in column refer to the target dataset shown in row 1. Rows 2, 3: GP posterior means of the fused/pre-transformed and fused/transformed datasets with covariance hyperparameters estimated. Rows 4 and 5: Window weights after min-phase reconstruction into time-domain and equalization-transform weights in absolute log-space. Row 6: SDRs (larger is more more accurate) of GP posterior means for original (+), fused/pre-transformed (*), and fused/transformed (x) datasets at test inputs and observations corresponding to target dataset.

datasets are larger than that of the pre-transformed ones in Fig. 3 across most frequency bands; SDRs may be compared against the control (GP interpolant \hat{f}_t with only the target dataset \mathcal{D}_t as evidence). Larger SDR discrepancies between median-plane HRTFs² than that of the horizontal-plane for transformed and control cases may indicate greater measurement sensitivities in the former directions. Fused dataset target plots {6, 5} and {6, 7} have the highest SDRs relative to the control. The equalization weights for each frequency appear continuous in log-space. The window weights exhibit periodicity similar to window functions in the Fourier domain. Moreover, both window and equalization weights for median and horizontal plane HRTFs from the same target \mathcal{D}_t are similar.

6. CONCLUSIONS

We have presented a joint spatial-frequency GP fusion method for combining common subject HRTF datasets via a set of linear transformations. Transformation filter weights are automatically learned for horizontal and median-plane “Club Fritz” HRTF measurements, which characterize inter-dataset measurement variances. This is verified in experiments where transformed cases produce larger SDRs than unfused cases relative to target datasets. Future work will consider non-linear transformations between HRTFs from different subjects for the personalization problem. The techniques developed may also find application in other problems where the same signal from different measurements must be compared.

²One horizontal-plane only dataset was omitted as a median-plane target.

7. ACKNOWLEDGMENT

We thank Dr. B. F. G. Katz and Dr. D. R. Begault for organizing the round-robin activity and hosting the club Fritz datasets.

8. REFERENCES

- [1] D. Zotkin, R. Duraiswami, and L. S. Davis, “Rendering localized spatial audio in a virtual auditory space,” *IEEE Transactions on Multimedia*, vol. 6, pp. 553–564, 2004.
- [2] B. F. G. Katz and D. R. Begault, “Round robin comparison of HRTF measurement system: preliminary results,” in *Proceedings of ICA*, 2007.
- [3] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Cambridge, Massachusetts: MIT Press, 2006.
- [4] Y. Luo, D. N. Zotkin, H. Daumé III, and R. Duraiswami, “Kernel regression for head-related transfer function interpolation and spectral extrema extraction,” in *ICASSP*, 2013.
- [5] Y. Saatci, “Scalable inference for structured Gaussian process models,” Ph.D. dissertation, University of Cambridge, 2011.
- [6] G. E. Uhlenbeck and L. S. Ornstein, “On the theory of Brownian motion,” *Phys. Rev.*, vol. 36, pp. 823–841, 1930.