GAUSSIAN PROCESS MODELS FOR HRTF BASED 3D SOUND LOCALIZATION

Yuancheng Luo, Dmitry N. Zotkin, Ramani Duraiswami

Department of Computer Science and Institute for Advanced Computer Studies
University of Maryland, College Park MD 20742 USA

ABSTRACT

The human ability to localize sound-source direction using just two receivers is a complex process of direction inference from spectral cues of sound arriving at the ears. While these cues can be described using the well-known head-related transfer function (HRTF) concept, it is unclear as to how densely HRTF must be sampled and whether a higher-order representation is employed in localization. We propose a class of binaural sound source localization models to answer these two questions. First, using the sound received by two ears, we derive several binaural features that are invariant to the sound source signal. Second, these are implicitly mapped to a high-dimensional *reproducing kernel Hilbert space* via a Gaussian process regression model for feature-direction tuples. Lastly, the features that are most relevant in the model are found via an efficient forward subset-selection method. Experimental results are shown for HRTFs belonging to the CIPIC database.

Index Terms— Gaussian process regression, head-related transfer function, source cancellation algorithm, subset selection

1. INTRODUCTION

Many animals posess a remarkable omnidirectional sound localization ability enabled by subconsciously processing subtle differences between the sounds received in two ears from a common source location. For humans, these differences are due to the incoming acoustic wave scattering off the listener's anatomic features (head, torso, pinna) before reaching the eardrum. The spectral ratio between the sounds recorded at the eardrum and at the center of the head in absence of the listener is known as the head-related transfer function (HRTF) [1]; it is specific to the individual's anthropometry and to the wave direction. HRTF contains important cues such as interaural time delay (ITD) and interaural level difference (ILD) [2] and HRTF use for virtual auditory synthesis provides life-like accuracy in sound localization [3, 4, 5].

A source localization algorithm employing only two receivers would be of interest in machine perception area (e.g. for event detection and localization [6]). However it is currently unknown how the mapping between binaural cues and source location is done in mammalian brain. In this paper, we investigate the sound source localization (SSL) problem in terms of binaural features derived from signals received by left / right ears in response to a static sound source (see section 2). Our features are not unlike the HRTF (in particular, they do not depend on source signal) and can be viewed as HRTF generalization. The feature representation's relevance to sound-source directions is explored within non-parametric regression frameworks such as Gaussian process regression (GPR) [7].

In the SSL problem, regression models between predictor (binaural features) and output (sound source directions) variables are learned. GPR is a non-parametric (number of parameters is proportional to the number of feature to direction inputs) *kernel method* that places weak assumptions on prior mean and covariance between output variables and is capable of automatic model-order selection and Bayesian inference. The predictor variables implicitly map to a reproducing kernel Hilbert space whose inner products are taken to be evaluations of a valid Mercer kernel or covariance function [8]. Moreover, Gaussian processes (GPs) generalize nearest-neighbor (NN) methods[1] as the former can infer outputs outside the training set; the predicted sound source directions (GP posterior mean function) are linear combinations of non-linear covariance evaluations between training and test feature inputs (see Fig. 1 and section 3).

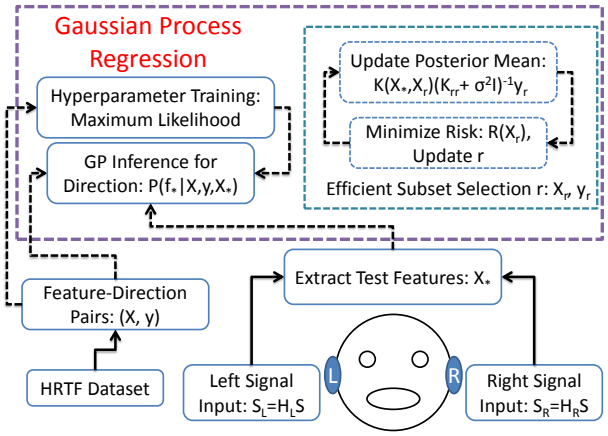

Fig. 1. GP training, inference, and subset-selection sub-systems.

The non-parametric aspect allows one to determine which input samples are most relevant to GP inference. We use a standard greedy forward-selection (GFS) method [9] that sequentially includes new feature-direction tuples (without replacement) in section 3.1. The ranking of these samples depends on an objective risk function to be minimized. We derive such a class of risk functions based on GP inference, which is efficiently computed (see section 3.2). Experimental results with the CIPIC HRTF database [10] show that the GP models are more accurate than least squares (LS) and NN methods (see section 4). We also show that subset selection via GFS results in much lower generalization GP error than random subset selection.

Related Work: In human and machine perception, a number of authors have attempted to use HRTF in SSL [11]. For example, in the matched filtering / source cancellation algorithms, the ratio between left and right ear signals [12] [13] [14] [15] is cross-correlated

Support of NSF award IIS-1117716 is gratefully acknowledged.

[1]Non-parametric classifier where each HRTF feature and measurement direction is a separate class exemplar and label.

Table 1. Derivation of feature variables X from signals S_L and S_R.

$\log\left(\left\|\frac{S_L}{S_R}\right\|+1\right)=\log\left(\frac{\|H_L\|}{\|H_R\|}+1\right)$	Log-magnitude ratio
$\angle\frac{S_L}{S_R}=\angle H_L-\angle H_R$	Phase difference
$\frac{\|S_L\|}{0.5(\|S_L\|+\|S_R\|)}=\frac{2\|H_L\|}{\|H_L\|+\|H_R\|}$	Avg magnitude ratio
$\{\|S_L\|,\|S_R\|\}=\{\|H_L\|,\|H_R\|\}$	Magnitude pair

with all left / right HRTF ratios over the sphere of directions and the direction maximizing correlation is presumed to be source direction. Other authors use cue mapping [16] [6] [17] using features such as ITD, ILD, interaural envelope difference, and notch frequency and classifiers such as linear regression, weighted kernel-NN, and self-organizing map. Statistical models have also been proposed; [18] derives a conditional probability map from ITD / ILD to infer direction via a maximum aposteriori estimator [18], and [19] learns a probabilistic affine regression model between interaural transfer functions and direction.

2. FEATURE REPRESENTATIONS

Using HRTF definition [1], we express the received left/right ear signals S_L, S_R as convolution of the source signal S with left/right HRTFs H_L, H_R for the source direction; in the frequency domain, convolution reduces to pointwise multiplication $S_L = H_L \circ S$ and $S_R = H_R \circ S$.

In Table 1, we derive several HRTF-like[2] features. The log-magnitude ratio is a percentage change in the relative signal loudness in left and right ears. The phase difference contains mostly ITD information related to source azimuth. The average magnitude ratio is a difference in contribution of left/right HRTF to the binaural average. The magnitude pair is essentially the left and right HRTFs *per se* concatenated. Note that each feature is frequency-dependent and is thus a vector with length equal to the number of frequency bins in Fourier transform; for CIPIC database with HRIR length of 200 samples, the vector length is 200 for the magnitude pair feature and 100 for the three others.

3. 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. In a parametric model, the problem is one of estimating a fixed set of parameters based on training data. When a parametric model is unknown, a common Bayesian approach for inference marginalizes over a distribution of latent function realizations $f(x)$ assumed to have generated the noisy observations in $y = f(x) + \epsilon$; noise term $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is zero centered with constant variance σ^2.

A GP f treats latent function realizations $f(x)$ as a collection of random variables where any finite subset sampled at inputs $X = [x_1, \ldots, x_N]$ is multivariate Gaussian $[f(x_1), \ldots, f(x_N)] \sim \mathcal{N}(\mu(x), K(X,X))$ and defined by a prior mean $\mu(X) \in \mathbb{R}^{N\times 1}$ and covariance or Gram matrix $K(X,X) \in \mathbb{R}^{N\times N}$ of pairwise covariance function evaluations between inputs $x \in X$. The prior mean $\mu(x)$ can be taken to be zero without loss of generality.

For N known training input pairs in X, y and N_* test pairs in $X_*, f_* = f(X_*)$, GP inference follows the conditioning of test outputs on the test inputs and training pairs, which is also a multivariate Gaussian $P(f_*|X, y, X_*) \sim \mathcal{N}(\bar{f}_*, \bar{\Sigma}_*)$ defined by posterior mean and covariance functions given by

$$\bar{f}_* = K_{f*}^T \hat{K}^{-1} y, \quad \bar{\Sigma}_* = K_{**} - K_{f*}^T \hat{K}^{-1} K_{f*}, \tag{1}$$

where $\hat{K} = K(X,X) + \sigma^2 I$ adjusts for the observation noise and $K_{f*} = K(X, X_*) \in \mathbb{R}^{N\times N_*}$ are pair-wise covariance evaluations between training and test inputs. In our SSL model, we treat the posterior mean $\bar{f}_*$ for inputs X_* in Eq. 1 as one of three predicted coordinates along the standard basis; three independent GPs are specified and trained on a common set of features X in Table 1 that map to single coordinate direction y normalized to be on the unit sphere. Each GP is initialized with the same covariance function whose hyperparmeters can be trained without cross-validation.

We specify the covariance function as products of identical class (ν) Matérn functions across each of the D input variables given by

$$\begin{aligned} &K_{\frac{1}{2}}(r,\ell) = e^{-\frac{r}{\ell}}, \quad K_{\frac{3}{2}}(r,\ell) = \left(1+\frac{\sqrt{3}r}{\ell}\right)e^{-\frac{\sqrt{3}r}{\ell}}, \\ &K_{\infty}(r,\ell) = e^{-\frac{r^2}{2\ell^2}}, \quad K(x,x') = \alpha^2 \prod_{k=1}^{D} K_\nu(|x_k - x'_k|, \ell_k), \end{aligned} \tag{2}$$

for distance r and hyperparameters α, ℓ_k. In the general case, a hyperparameter Θ can be optimized via maximum log-marginal data likelihood function with its gradient given by

$$\begin{aligned} \log p(y|X) &= -\frac{1}{2}\left(\log|\hat{K}| + y^T\hat{K}^{-1}y + N\log(2\pi)\right), \\ \frac{\partial \log p(y|X)}{\partial \Theta_i} &= -\frac{1}{2}\left(\mathrm{tr}\left(\hat{K}^{-1}P\right) - y^T\hat{K}^{-1}P\hat{K}^{-1}y\right), \end{aligned} \tag{3}$$

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

3.1. Greedy Forward-Selection

Feature selection in a non-parametric setting such as GPR finds a subset of feature-direction training samples that best generalizes a test set. The main advantage of a greedy subset-selection heuristic is the reduction of a general NP-hard search problem to the one with the cost quadratic in terms of number of the objective risk function R evaluations. Algorithm 1 summarizes this approach; for subset r at round t, inputs $X_{\hat{r}\notin r}$ are ranked by the risk function $R(X_{r\cup\hat{r}})$ and one with the lowest rank is incorporated into the GP model and is discarded from consideration for all future rounds.

Algorithm 1 Greedy Forward Selection (**GFS**)

Require: Training inputs X, y, and risk function $R(X)$.
1: $r \leftarrow \emptyset$ \\ Initial empty subset at round $t = 0$
2: **for** $t = 1$ to N **do**
3: $r \leftarrow \{r, \arg\min_{\hat{r}\notin r} R(X_{r\cup\hat{r}})\}$ \\ Minimize risk
4: **end for**
5: **return** r

Adding a new input pair $(x_{\hat{r}}, y_{\hat{r}})$ to a GP model at round t is efficient due to the proposed recurrence relations between Gram matrices before and after the union of subset r with $\hat{r}$ in $\breve{r} = r\cup\hat{r}$ given by

$$\begin{aligned} K_{(\breve{r})} &= \begin{bmatrix} K_{(r)} & k_{r\hat{r}} \\ k_{r\hat{r}}^T & k_{\hat{r}\hat{r}} \end{bmatrix} = \begin{bmatrix} K_{(r)} & 0 \\ 0 & 1 \end{bmatrix} - uu^T + vv^T, \\ k_{r\hat{r}} &= K(X_r, X_{\hat{r}}), \quad k_{\hat{r}\hat{r}} = K(X_{\hat{r}}, X_{\hat{r}}) + \sigma^2, \end{aligned} \tag{4}$$

[2] Similarly to HRTF, they do not depend on the source signal S; hence, in the remainder of the paper we compute them directly from HRTF.

where vectors $u = \sqrt{\frac{||w||}{2}}\left(\frac{w}{||w||} + e_t\right)$, $v = \sqrt{\frac{||w||}{2}}\left(\frac{w}{||w||} - e_t\right)$, $w = \left[-k_{r\hat{r}}^T, \frac{1-k_{\hat{r}\hat{r}}}{2}\right]^T$, and e_t is the t^{th} column of the identity matrix. The inverse covariance matrix follows the modified *Woodbury* formulation [20] given by

$$\begin{aligned} K_{(\check{r})}^{-1} &= \bar{K}^{-1} + d_u \bar{u}\bar{u}^T - d_v \bar{v}\bar{v}^T, \quad \bar{K}^{-1} = \begin{bmatrix} K_{(r)}^{-1} & 0 \\ 0 & 1 \end{bmatrix}, \\ \bar{u} &= \bar{K}^{-1}u, \quad d_u = (1 - < \bar{u}, u >)^{-1}, \\ \bar{v} &= \left(\bar{K}^{-1} + d_u \bar{u}\bar{u}^T\right) v, \quad d_v = (1 + < \bar{v}, v >)^{-1}; \end{aligned} \tag{5}$$

as such, only two rank-1 updates are required. The new log-determinant is given by $\log\left|K_{(\check{r})}\right| = \log\left|\bar{K}\right| - \log d_u d_v$. For fixed test inputs, the posterior mean remains a matrix-vector product and the posterior variances are sums of diagonals given by

$$\begin{aligned} \bar{f}_{*\check{r}} &= K_{*\check{r}} K_{(\check{r})}^{-1} y_{\check{r}}, \quad s_u = K_{*\check{r}}\bar{u}, \quad s_v = K_{*\check{r}}\bar{v}, \\ \textbf{diag}\left(\bar{\Sigma}_{*\check{r}}\right) &= \textbf{diag}\left(\bar{\Sigma}_{*r} + k_{*\hat{r}}k_{*\hat{r}}^T + d_u s_u s_u^T - d_v s_v s_v^T\right), \end{aligned} \tag{6}$$

where matrix $K_{*\check{r}} = K(X_*, X_{\check{r}})$. Updating the GP prior and posterior models require $\mathrm{O}\left(t^2\right)$ and $\mathrm{O}\left(N_* t\right)$ operations respectively.

3.2. Risk Function Criterions

One class of risk functions R is the L^2 Euclidean distance between two functions evaluated at a possibly infinite sized set of test inputs X_*. The L^2 distance between two GP posterior mean functions $\bar{f}_a = K_{*a}\hat{K}_a^{-1}y_a$ and $\bar{f}_b = K_{*b}\hat{K}_b^{-1}y_b$ is analytic assuming that prior mean $m(x) = 0$ and that the covariance function belongs to Matérn class (Eq. 2). Then, the squared errors over the set of test inputs X_* is given by

$$\begin{aligned} L_{X_*}^2\left(\bar{f}_a, \bar{f}_b\right) &= \sum_{x_* \in X_*} \left(\bar{f}_a - \bar{f}_b\right)^2 \\ &= z_a^T Q_{aa} z_a - 2 z_a^T Q_{ab} z_b + z_b^T Q_{bb} z_b, \end{aligned} \tag{7}$$

where vectors $z_a = \hat{K}_a^{-1}y_a \in \mathbb{R}^{N_a}$, $z_b = \hat{K}_b^{-1}y_b \in \mathbb{R}^{N_b}$ are computed over training data. If the set of test inputs X_* is finite, then matrices $Q_{aa} = \sum_{x_* \in X_*} K_{a*}K_{*a} \in \mathbb{R}^{N_a \times N_a}$, $Q_{ab} = \sum_{x_* \in X_*} K_{a*}K_{*b} \in \mathbb{R}^{N_a \times N_b}$, and $Q_{bb} = \sum_{x_* \in X_*} K_{b*}K_{*b} \in \mathbb{R}^{N_b \times N_b}$ are sums of outerproducts whose i, j^{th} entry are products of Matérn class covariance functions in Eq. 2. If the set of test inputs $X_* = (-\infty, \infty)$ is the full input domain, then matrices $Q_{aa} = \int_{-\infty}^{\infty} K_{a*}K_{*a}dx_* \in \mathbb{R}^{N_a \times N_a}$, $Q_{ab} = \int_{-\infty}^{\infty} K_{a*}K_{*b}dx_* \in \mathbb{R}^{N_a \times N_b}$, and $Q_{bb} = \int_{-\infty}^{\infty} K_{b*}K_{*b}dx_* \in \mathbb{R}^{N_b \times N_b}$ contain improper integral entries. For a valid distance measure, the posterior mean functions converge to identical zero-mean priors at the limits $x_{*k} \to \pm\infty$ and the improper integrals of the form $Q_{a_i b_j} = \prod_{k=1}^{D} F_{\nu ijk}$ given by

$$F_{\nu ijk} = \int_{-\infty}^{\infty} K_\nu(|x_{a_i k} - x_{*k}|, \ell_{ak}) K_\nu(|x_{b_j k} - x_{*k}|, \ell_{bk}) dx_{*k}, \tag{8}$$

are shown to be finite (see Appendix Eq. 10).

Several combinations of the L^2 distance are shown in Table 2: The prediction error is taken at known feature-direction tuples. The generalized error is evaluated at a finite set of test inputs (out-of-sample) between GPs evidenced on the subset-selected and full set of feature to direction pairs. The normalized error ("frequentist") risk is integrated over the entire input domain X_* with uniform probability distribution where $\|f\| = \sqrt{\int_{-\infty}^{\infty} f(x)^2 dx}$. Note that computing the risk function between successive rounds t is efficient as the posterior mean function $\bar{f}_{(\check{r})} = K_{(*\check{r})}K_{(\check{r})}^{-1}y_{\check{r}}$ need only rank-1 updates via Eq. 5. The associated matrix $Q_{\check{r}X}$ in Eq. 7 evaluated between subset $\check{r}$ and the full input set X is a sub-matrix of the pre-computed matrix Q_{XX}. Criterion functions such as information gain are not considered as posterior covariance related functions (inverse, determinant) are expensive to compute or may be intractable.

Table 2. Combination of risk functions R

$L_X^2\left(\bar{f}_{(\check{r})}, y\right)$	Prediction error at known y
$L_{X_*}^2\left(\bar{f}_{(\check{r})}, \bar{f}_{(X)}\right)$	Generalized error at any X_*
$L_{(-\infty,\infty)}^2\left(\frac{\bar{f}_{(\check{r})}}{\|\bar{f}_{(\check{r})}\|}, \frac{\bar{f}_{(X)}}{\|\bar{f}_{(X)}\|}\right)$	Normalized error over domain

4. EXPERIMENT RESULTS

For training data, we use the left and right ear HRTFs from the CIPIC [10] database belonging to subject 156. Our error metric follows the angular separation distance between two directions **u, u'** given by

$$\text{dist}\left(\mathbf{u}, \mathbf{u'}\right) = \cos^{-1}\frac{< \mathbf{u}, \mathbf{u'} >}{||\mathbf{u}||||\mathbf{u'}||}, \quad \mathbf{u}, \mathbf{u'} \in \mathbb{R}^3. \tag{9}$$

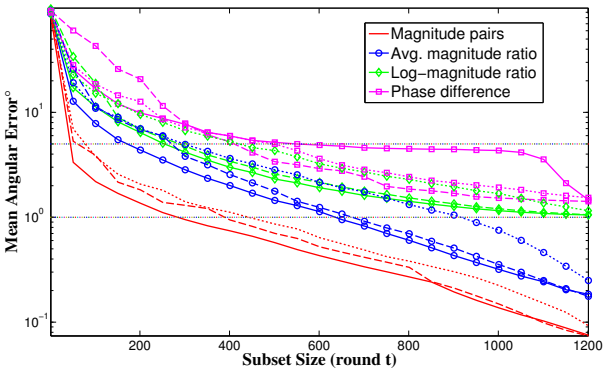

Fig. 2. Mean angular errors for subset-selected GPs with covariance K_∞ are shown across feature types and rounds. Both risk functions for prediction error (solid) and normalized error (dashed) show better generalization than randomized (dotted) subsets. Intercepts with horizontal lines indicate subset sizes at 5° and 1° errors.

Cross-Validation: GP models are trained on a randomized data subset[3] and predict spatial directions on the full dataset; GPs evidenced on the randomized subset infer directions at test inputs $X_* = X$ via Eq. 1. For a baseline, we compare against NN[4] and OLS[5] methods. The mean angular separation between predicted and reference directions in Eq. 9 are computed along all, horizontal plane, and median plane directions in Table 3; minimum errors per feature and subset of directions are in bold. Non-parametric methods

[3] Training is done on $1/3$ of all available data; the training subset contains $1250/3 = 417$ feature-direction tuples. Hyperparameters are trained for 50 iterations using resilient backpropagation [21].

[4] Euclidean distance between full and training feature sets.

[5] $y = X\beta$ for input features X and parameters β.

Table 3. Mean angular errors in degrees (full, h-plane, m-plane)

	Mag. pair	Avg. ratio	Log ratio	Phase diff.
OLS	5.4, 4.7, 5.3	22, 20, 18	29, 32, 25	27, 21, 22
NN	3.9, 4.7, 4.2	7.9, 9.2, 11	9.2, 10, 11	20, 22, 28
GPR $K_{1/2}$	1.8, 1.7, 2.5	7, 7.4, 12	7.2, **6.8**, **8.9**	12, 12, 15
GPR $K_{3/2}$	1.4, 1.5, 2.2	4.8, 5.5, **8.2**	7.5, 9.5, 11	11, 14, 13
GPR K_{∞}	**1.3**, **1.3**, **1.6**	**4.8**, **4.8**, 8.8	**6.3**, 12, 10	**6.3**, **5.2**, **13**

as NN and GPR outperform parametric methods as OLS across all feature types. The left and right HRTF magnitude pairs and average magnitude ratio features give the lowest errors; their results are shown in the first and the third plots in Fig. 3. OLS results suggest that log-ratios are oversensitive predictors of change in localization direction, whereas NN results suggest that using phase difference / ITD is insufficient for localizing in elevation.

Greedy Forward-Selection: The relevant input samples are found using the GFS Algorithm 1 using varying subset size. At round t, the updated GP model infers directions along the full set of input features. The angular separation errors w.r.t. reference directions are computed in Fig. 2. The crossover points at 5° mean angular error show that magnitude pair and average magnitude ratios require about 50 and 150 training samples for human-level accuracy. Both risk function criterions outperform randomized subset selection in all but the phase difference features. The improvement over the randomized set is apparent in the second and fourth plots of Fig. 3 where directions further from the vertical plane are more accurately localized.

5. CONCLUSIONS

We developed a robust SSL method using source-independent, HRTF-like features and GP models. Our method generalizes NN-based techniques and is more accurate due to efficient model selection; both GP hyperparameters and most relevant training samples were automatically learned. Experimental results have shown that accurate localization over the full sphere is possible using only a small fraction of the typically-sampled HRTF data points when average magnitude ratio features are used. We leave extensions to SSL in reverberant conditions using linear combinations of S_L, S_R signals for future work.

A. MATÉRN PRODUCT INTEGRALS

$$F_{\frac{1}{2}ijk} = \left(\ell_{ak} e^{\frac{-\left|x_{a_i k} - x_{b_j k}\right|}{\ell_{ak}}} - \ell_{bk} e^{\frac{-\left|x_{a_i k} - x_{b_j k}\right|}{\ell_{bk}}} \right) \frac{2\ell_{ak}\ell_{bk}}{\ell_{ak}^2 + \ell_{bk}^2},$$

$$F_{\frac{3}{2}ijk} = \left(\ell_{ak}^2 (\ell_{ak} - \beta\ell_{bk} - \alpha) e^{\frac{-\sqrt{3}\left|x_{a_i k} - x_{b_j k}\right|}{\ell_{ak}}} \right.$$

$$\left. + \ell_{bk}^2 (\ell_{bk} + \beta\ell_{ak} - \alpha) e^{\frac{-\sqrt{3}\left|x_{a_i k} - x_{b_j k}\right|}{\ell_{bk}}} \right) \frac{4\ell_{ak}\ell_{bk}}{\sqrt{3}\left(\ell_{ak}^2 - \ell_{bk}^2\right)^2},$$

$$\alpha = -\sqrt{3}\left|x_{a_i k} - x_{b_j k}\right|, \quad \beta = \frac{4\ell_{ak}\ell_{bk}}{\ell_{ak}^2 - \ell_{bk}^2},$$

$$F_{\infty ijk} = e^{-\frac{\left(x_{a_i k} - x_{b_j k}\right)^2}{2\left(\ell_{ak}^2 + \ell_{bk}^2\right)}} \frac{\ell_{ak}\ell_{bk}\sqrt{2\pi}}{\sqrt{\ell_{ak}^2 + \ell_{bk}^2}}. \tag{10}$$

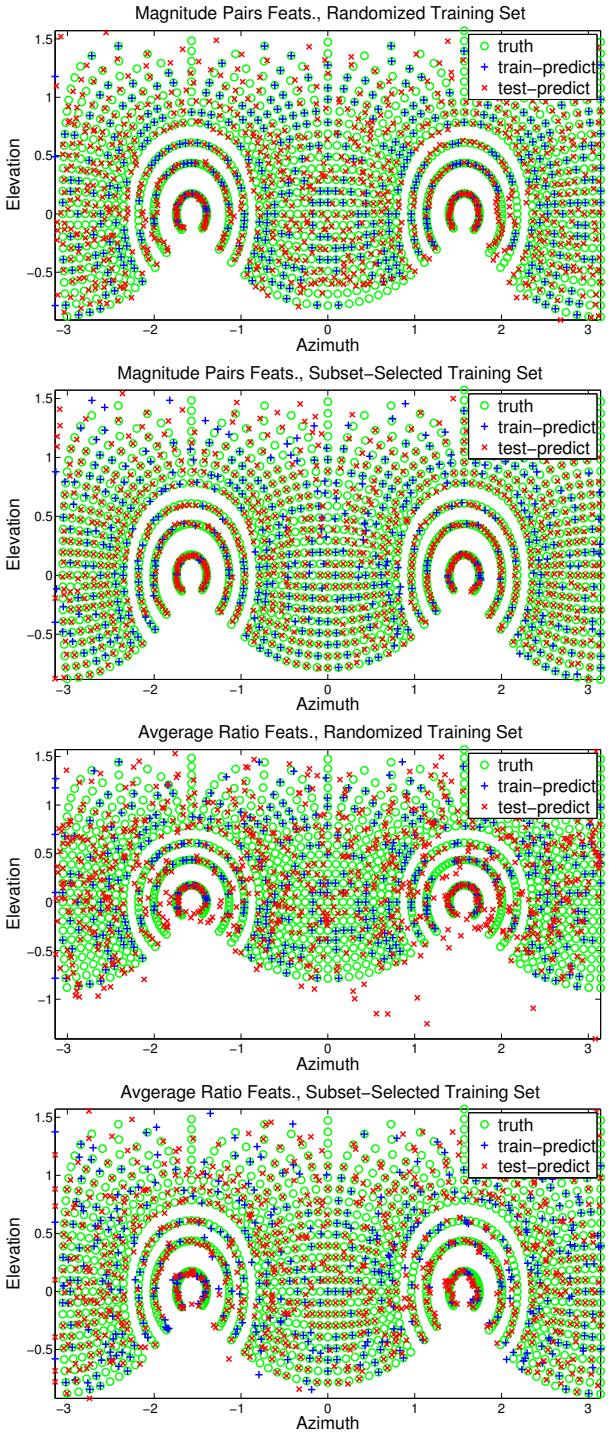

Fig. 3. Mercator projections of GP (covariance K_{∞}) predicted directions evidenced with randomized and subset-selected inputs (prediction error risk function R in Table. 2) are shown.

6. REFERENCES

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

[2] C. Cheng and G. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," in *Audio Engineering Society Convention 107*, 1999.

[3] A. Kulkarni and H. Colburn, "Role of spectral detail in sound-source localization," *Nature*, vol. 396, no. 6713, pp. 747–749, 1998.

[4] E. Wenzel, M. Arruda, D. Kistler, and F. Wightman, "Localization using nonindividualized head-related transfer functions," *JASA*, vol. 94, pp. 111, 1993.

[5] G. Romigh, D. Brungart, R. Stern, and B. Simpson, "The role of spatial detail in sound-source localization: Impact on HRTF modeling and personalization.," in *Proceedings of Meetings on Acoustics*, 2013, vol. 19, p. 050170.

[6] J. Hornstein, M. Lopes, J. Santos-Victor, and F. Lacerda, "Sound localization for humanoid robots-building audio-motor maps based on the HRTF," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. IEEE, 2006, pp. 1170–1176.

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

[8] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *Neural Networks, IEEE Transactions on*, vol. 12, no. 2, pp. 181–201, 2001.

[9] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[10] 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*, New Paltz, NY, 2001, pp. 99–102.

[11] M. Rothbucher, D. Kronmüller, M. Durkovic, T. Habigt, and K. Diepold, "HRTF sound localization," 2011.

[12] F. Keyrouz, K. Diepold, and S. Keyrouz, "High performance 3D sound localization for surveillance applications," in *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*. IEEE, 2007, pp. 563–566.

[13] F. Keyrouz, "Humanoid hearing: A novel three-dimensional approach," in *Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 214–219.

[14] F. Keyrouz and K. Diepold, "An enhanced binaural 3D sound localization algorithm," in *Signal Processing and Information Technology, 2006 IEEE International Symposium on*. IEEE, 2006, pp. 662–665.

[15] A. Pourmohammad and S. Ahadi, "TDE-ILD-HRTF-Based 3D entire-space sound source localization using only three microphones and source counting," in *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on*. IEEE, 2011, pp. 1–6.

[16] T. Rodemann, M. Heckmann, F. Joublin, C. Goerick, and B. Scholling, "Real-time sound localization with a binaural head-system using a biologically-inspired cue-triple mapping," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. IEEE, 2006, pp. 860–865.

[17] H. Nakashima and T. Mukai, "3D sound source localization system based on learning of binaural hearing," in *Systems, Man and Cybernetics, 2005 IEEE International Conference on*. IEEE, 2005, vol. 4, pp. 3534–3539.

[18] V. Willert, J. Eggert, J. Adamy, R. Stahl, and E. Körner, "A probabilistic model for binaural sound localization," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, vol. 36, no. 5, pp. 1, 2006.

[19] A. Deleforge and R. Horaud, "2D sound-source localization on the binaural manifold," in *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on*. IEEE, 2012, pp. 1–6.

[20] R. Saigal, "On the inverse of a matrix with several rank one updates," Tech. Rep., University of Michigan Ann Arbor, 1993.

[21] M. Riedmiller, "RPROP: Description and implementation details," Tech. Rep., University of Karlsruhe, 1994.