

# Handling Missing Data with Variational Bayesian Estimation of ICA

**Kwokleung Chan**

and **Te-Won Lee**

*Computational Neurobiology Laboratory*

*The Salk Institute*

*10010 North Torrey Pines Road La Jolla, CA 92037, USA*

KWCHAN@SALK.EDU

TEWON@SALK.EDU

**John M. Allman**

*Division of Biology 216-76*

*California Institute of Technology*

*1200 E. California Blvd.*

*Pasadena, CA 91125*

CEBUS@CALTECH.EDU

**Terrence J. Sejnowski**

*Computational Neurobiology Laboratory*

*The Salk Institute*

*10010 North Torrey Pines Road*

*La Jolla, CA 92037, USA*

TERRY@SALK.EDU

## Abstract

Missing data is common in real-world datasets and is a problem for many estimation techniques. We have developed a variational Bayesian method to perform Independent Component Analysis (ICA) on high-dimensional data containing missing entries. Missing data are handled naturally in the Bayesian framework by integrating out from the generative density model. Modeling the distributions of the independent sources with mixture of Gaussians allows sources to be estimated with different kurtosis and skewness. Unlike the maximum likelihood approach, the variational Bayesian method automatically determines the dimensionality of the data and yields an accurate density model for the observed data without overfitting problems. An example will be given of an application to primate brain area data set containing many missing measurements.