COGNITIVE AND NEURAL SYSTEMS, BOSTON UNIVERSITY
Boston, Massachusetts

Latent Variable Framework for Modeling and Separating Single-Channel Acoustic Sources

Committee
Chair: Prof. Daniel Bullock
Readers: Prof. Barbara Shinn-Cunningham
Dr. Paris Smaragdis
Prof. Frank Guenther
Reviewers: Dr. Bhiksha Raj
Prof. Eric Schwartz

Madhusudana Shashanka
shashanka@cns.bu.edu

17 August 2007
Dissertation Defense

Outline

- Introduction
- Time-Frequency Structure } Background
- Latent Variable Decomposition:
  A Probabilistic Framework
- Sparse Overcomplete Decomposition
- Conclusions

Contributions

Introduction

The achievements of the ear are indeed fabulous. While I am writing, my elder son rattles the fire rake in the stove, the infant babbles contentedly in his baby carriage, the church clock strikes the hour, …

… In the vibrations of air striking my ear, all these sounds are superimposed into a single extremely complex stream of pressure waves. Without doubt the achievements of the ear are greater than those of the eye.

Wolfgang Metzger, in *Gesetze des Sehens* (1953)

Abridged in English and quoted by Reinier Plomp (2002)

Cocktail Party Effect

(*Cocktail Party* by SLAW, Maniscalco Gallery. From slides of Prof. Shinn-Cunningham, ARO 2006)

Colin Cherry (1953)

Our ability to follow one speaker in the presence of other sounds.

The auditory system separates the input into distinct *auditory objects.*

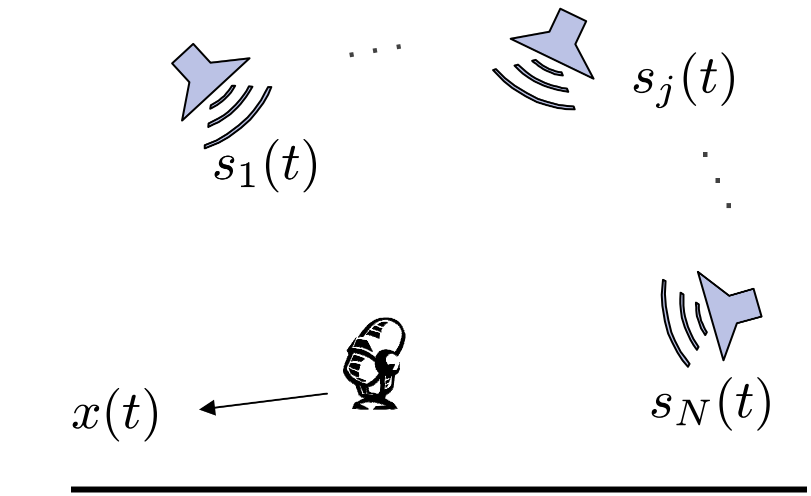
Challenging problem from a computational perspective.

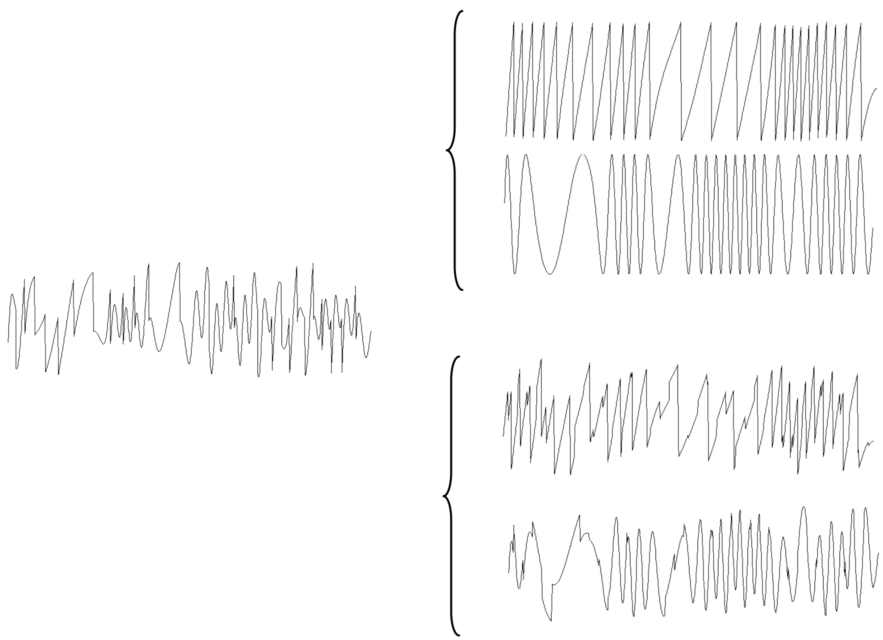
Cocktail Party Effect

- Fundamental questions
  - How does the brain solve it?
  - Is it possible to build a machine capable of solving it in a satisfactory manner?
    - Need not mimic the brain

- Two cases
  - Multi-channel (Human auditory system is an example with two sensors)
  - Single-Channel

Source Separation: Formulation





- $$x(t) = \sum_{j=1}^{N} s_j(t)$$

- Given just $x(t)$, how to separate sources $s_j(t)$?

- Problem: Indeterminacy
  - Multiple ways in which source signals can be reconstructed from the available information

Source Separation: Approaches

- *Exact* solutions not possible, but can *approximate*
  - by utilizing information about the problem
- Psychoacoustically/Biologically inspired approach
  - Understand how the auditory system solves the problem
  - Utilize the insights gained (rules and heuristics) in the artificial system
- Engineering approach
  - Utilize probability and signal processing theories to take advantage of known or hypothesized structure/statistics of the source signals and/or the mixing process

Source Separation: Approaches

- Psychoacoustically inspired approach
  - Seminal work of Bregman (1990) - *Auditory Scene Analysis (ASA)*
  - *Computational Auditory Scene Analysis (CASA)*
  - Computational implementations of the views outlined by Bregman (Rosenthal and Okuno, 1998)
  - Limitations: reconcile subjective concepts (e.g. “similarity”, “continuity”) with strictly deterministic computational platforms?
  - Difficulty incorporating statistical information

- Engineering approach
  - Most work has focused on multi-channel signals
  - Blind Source Separation: Beamforming and ICA
  - Unsuitable for single-channel signals

Source Separation: This Work

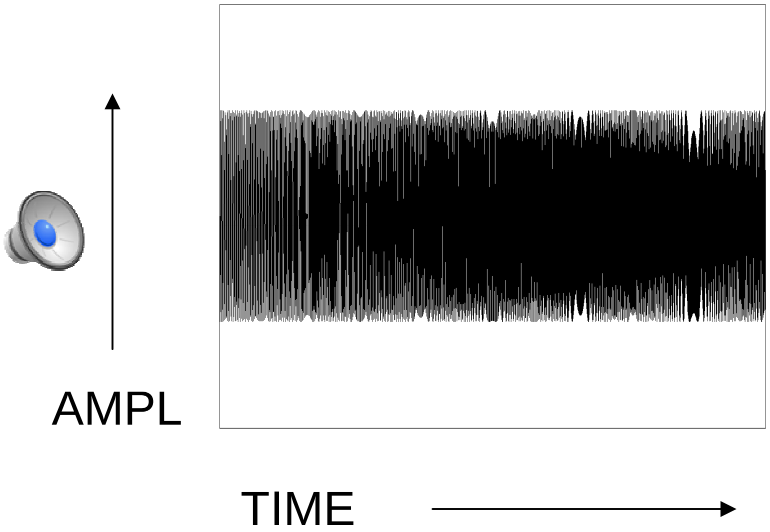
- We take a machine learning approach in a supervised setting
  - Assumption: One or more sources present in the mixture are "known"
  - Analyze the sample waveforms of the known sources and extract characteristics unique to each one
  - Utilize the learned *information* for source separation and other applications

- Focus on developing a probabilistic framework for modeling single-channel sounds
  - Computational perspective, goal not to explain human auditory processing
  - Provide a framework grounded in theory that allows principled extensions
  - Aim is not just to build *a* particular separation system

Outline

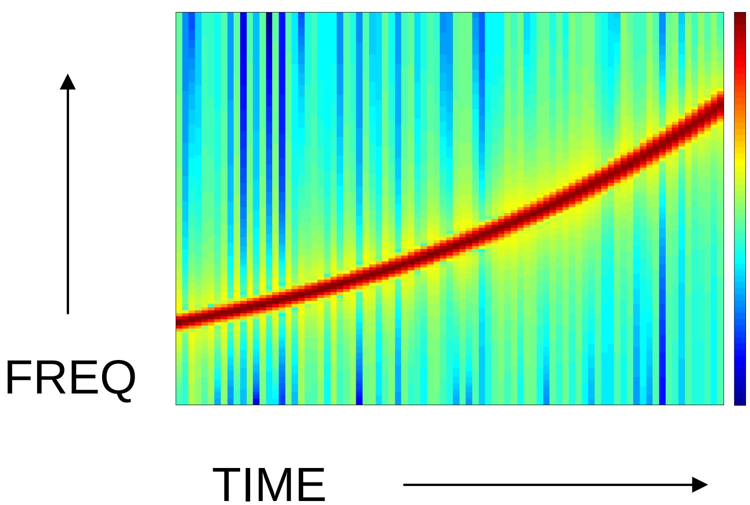
- Introduction
- Time-Frequency Structure
  - We need a representation of audio to proceed
- Latent Variable Decomposition:
  A Probabilistic Framework
- Sparse Overcomplete Decomposition
- Conclusions

Representation



Time-domain representation

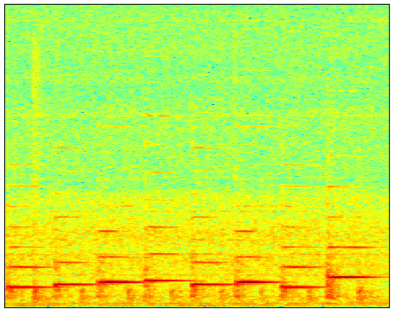
Sampled waveform: each sample represents the sound pressure level at a particular time instant.



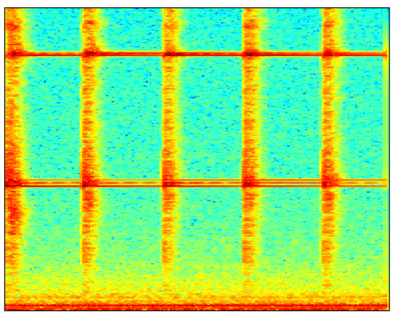
Time-Frequency representation

TF representation shows energy in TF bins explicitly showing the variation along time and frequency.

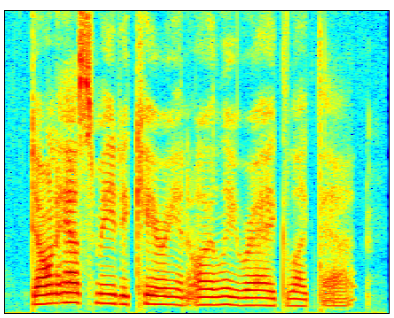
TF Representations



Piano



Cymbals

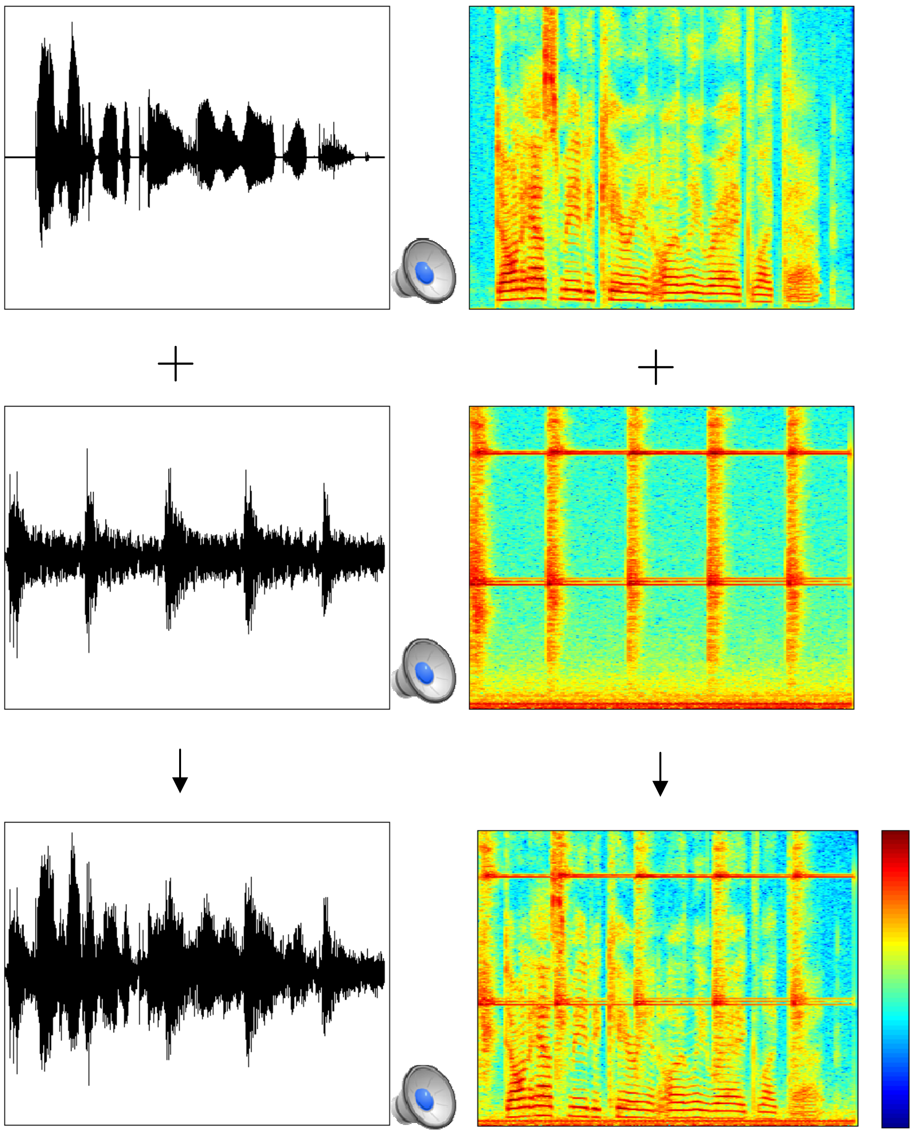


Female Speaker

Short Time Fourier Transform (STFT; Gabor, 1946)

- time-frames: successive fixed-width snippets of the waveform (windowed and overlapping)
- *Spectrogram*: Fourier transforms of all time slices. The result for a given time slice is a *spectral vector*.
- Other TF representations possible (different filter banks): only STFT considered in this work
    - Constant-Q (Brown, 1991)
    - Gammatone (Patterson et al. 1995)
    - Gamma-chirp (Irino and Patterson, 1997)
    - TF distributions (Laughlin et al. 1994)

Magnitude Spectrograms



Short Time Fourier Transform (STFT; Gabor, 1946)

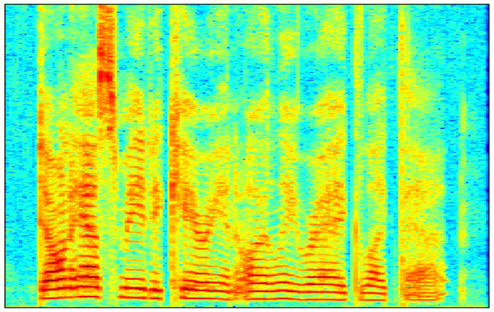
- Magnitude spectrograms: TF entries represent energy-like quantities that can be approximated to add additively in case of sound mixtures

- Phase information is ignored. Enough information present in the magnitude spectrogram, simple test:

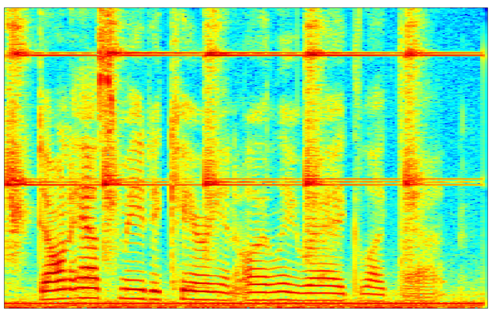
  - Speech with cymbals phase

  - Cymbals with piano phase

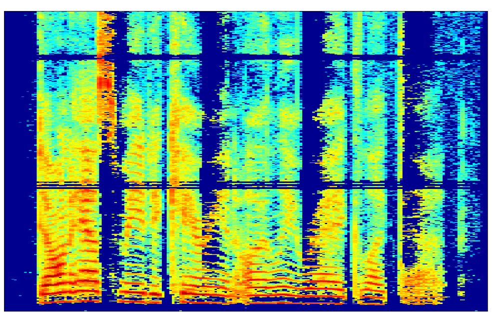
TF Masks



Target



Mixture

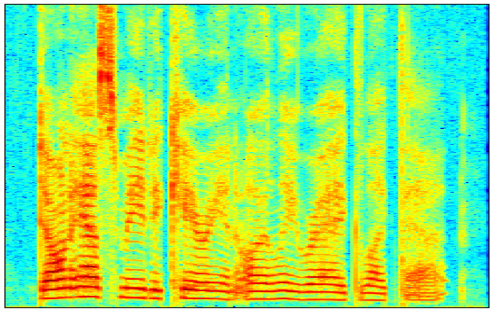


“Masked” Mixture

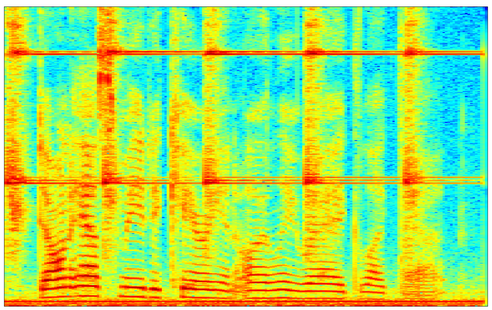
Time-Frequency Masks

- Popular in the CASA literature
- Assign higher weight to areas of the spectrogram where target is dominant
- Intuition: dominant source masks the energy of weaker ones in any TF bin, thus only such “dominant” TF bins are sufficient for reconstruction
- Reformulate the problem – goal is to estimate the TF mask (Ideal Binary Mask; Wang, 2005)
- Utilize cues like harmonicity, F0 continuity, common onsets/offsets etc.:
  - Synchrony strand (Cooke, 1991)
  - TF Maps (Brown and Cooke, 1994)
  - Correlograms (Weintraub, 1985; Slaney and Lyon, 1990)

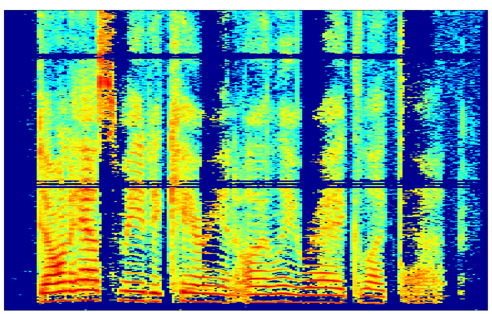
TF Masks



Target



Mixture



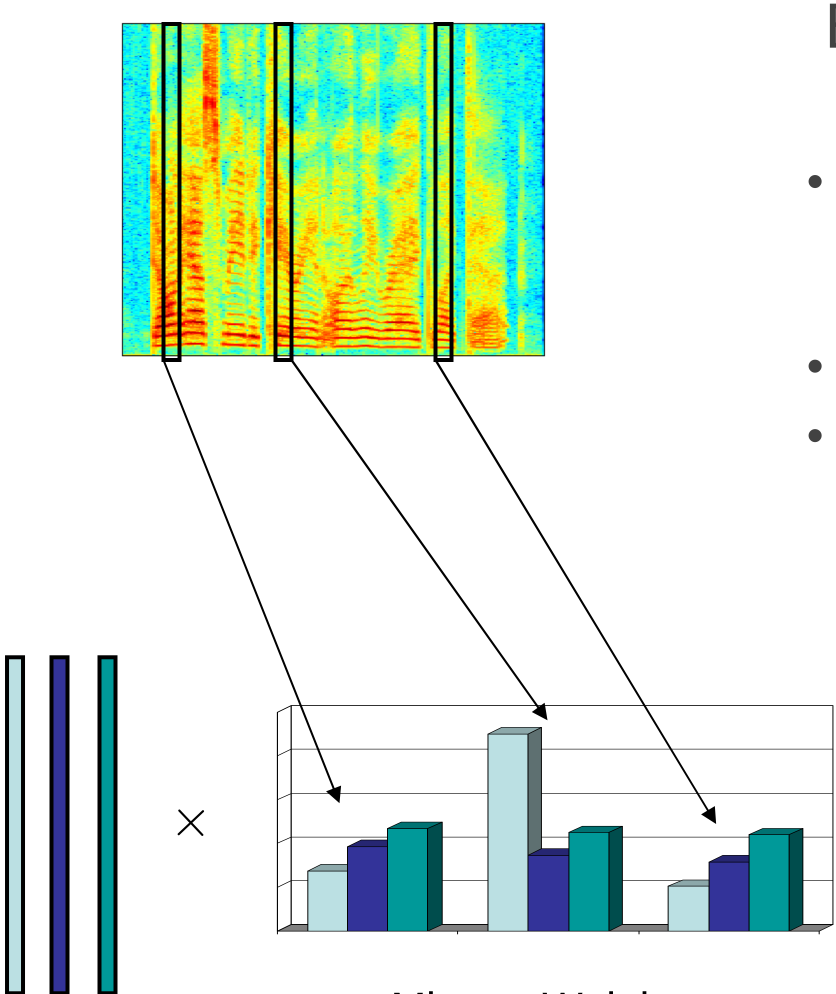
“Masked” Mixture

Time-Frequency Masks: Limitations

- Implementation of “fuzzy” rules and heuristics from ASA; ad-hoc methods, difficulty incorporating statistical information (Roweis, 2000)

- Assumption: energy sparsely distributed i.e. different sources are disjoint in their spectro-temporal content (Yilmaz and Rickard, 2004)
  - performs well only on mixtures that exhibit well-defined regions in the TF plane corresponding to the various sources (van der Kouwe *et al.* 2001)

Basis Decomposition Methods



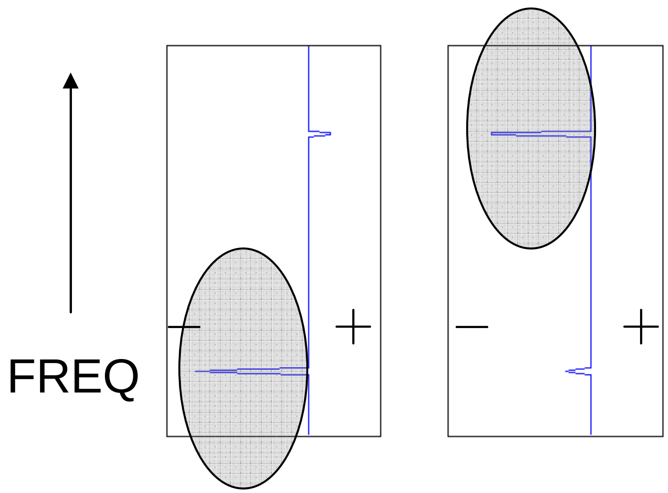
Basis Decomposition

- Idea: observed data vector can be expressed as a linear combination of a set of "basis components"
- Data vectors → spectral vectors
- Intuition: every source exhibits characteristic structure that can be captured by a finite set of components

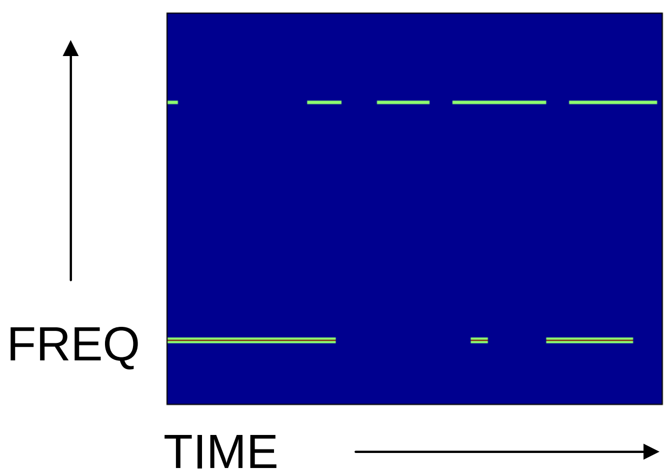
$$\mathbf{v}_t = \sum_{k=1}^{K} h_{kt} \mathbf{w}_k$$

$$\mathbf{V} = \mathbf{W}\mathbf{H}$$

Basis Decomposition Methods



PCA Components

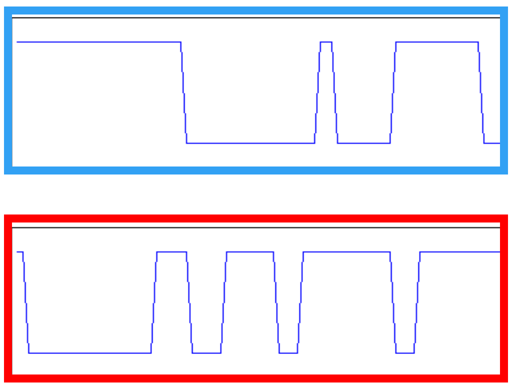


Basis Decomposition Methods

- Many Matrix Factorization methods available, e.g. PCA, ICA
- Toy example: PCA components can have negative values
- But spectrogram values are positive – interpretation?

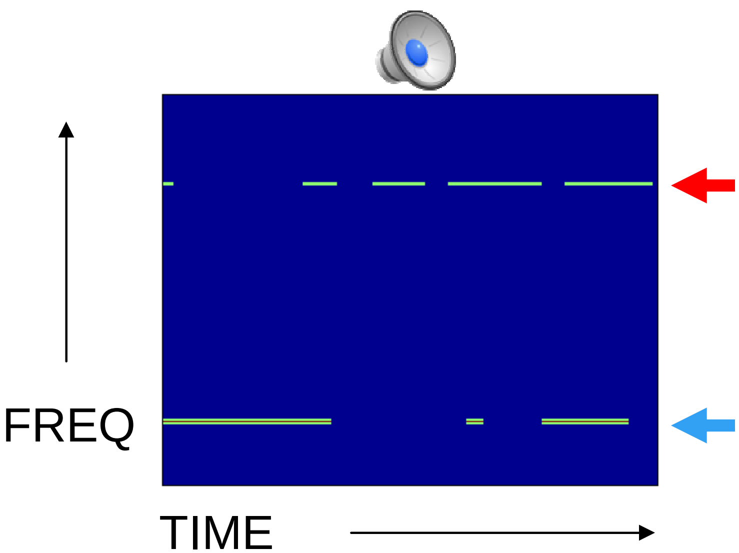
Basis Decomposition Methods

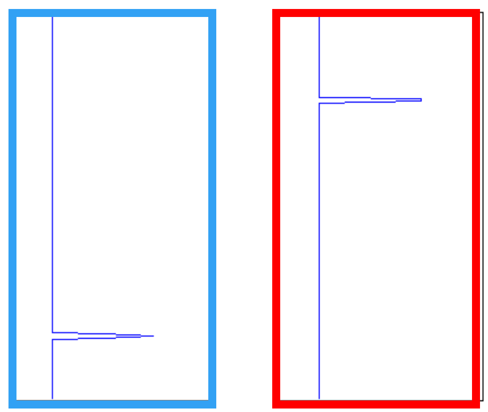
Mixture Weights
$\mathbf{H}$



Basis Decomposition Methods

- Non-negative Matrix Factorization (Lee and Seung, 1999)
- Explicitly enforces non-negativity on both the factored matrices
- Useful for analyzing spectrograms (Smaragdis, 2004, Virtanen, 2006)
- Issues
  - Can't incorporate prior biases
  - Restricted to 2D representations





$\mathbf{W}$ Basis Components

Outline

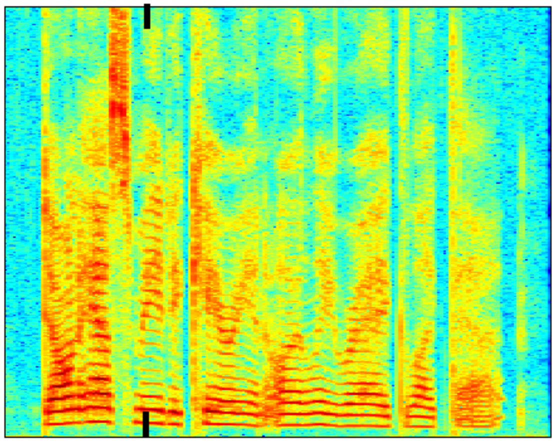
- Introduction
- Time-Frequency Structure
- Latent Variable Decomposition: Probabilistic Framework
  - Our alternate approach: Latent variable decomposition treating spectrograms as histograms
- Sparse Overcomplete Decomposition
- Conclusions

Latent Variables

- Widely used in social and behavioral sciences
  - Traced back to Spearman (1904), factor analytic models for Intelligence Testing

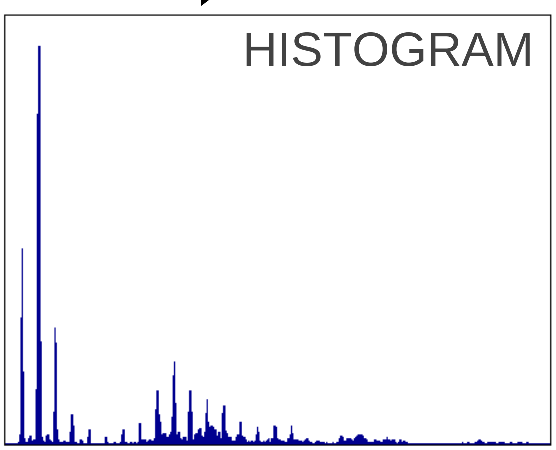
- Latent Class Models (Lazarsfeld and Henry, 1968)
  - Principle of local independence (or the *common cause criterion*)
  - If a latent variable underlies a number of observed variables, the observed variables conditioned on the latent variable should be independent

Spectrograms as Histograms



f

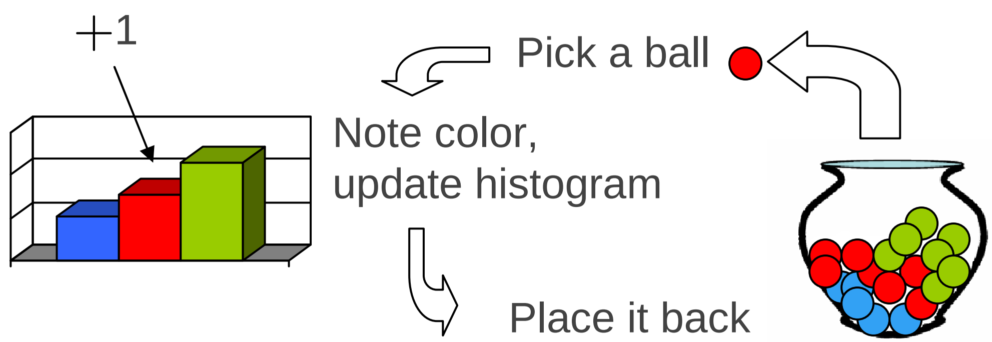
FRAME t



f

Generative Model

- Spectral vectors – energy at various frequency bins
- Histograms of multiple draws from a frame-specific multinomial distribution over frequencies
- Each draw → "a quantum of energy"

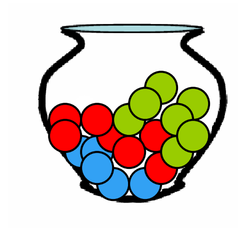


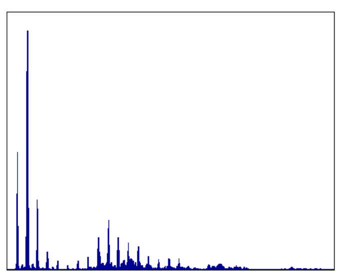
$P_t(f)$

Multinomial Distribution underlying the *t*-th spectral vector

Model

$$P_t(f)$$

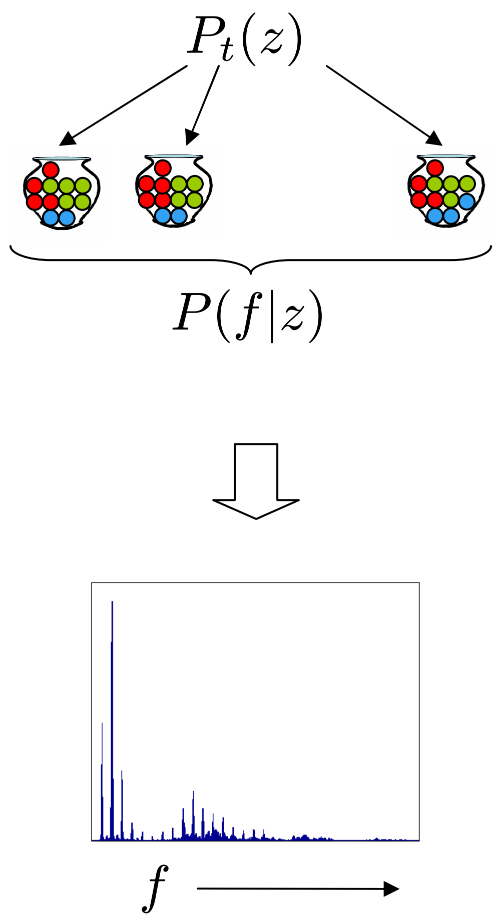




f

Model

$$P_t(f) = \sum_z P(f|z) P_t(z)$$



Generative Model

- Mixture Multinomial
- Procedure
  - Pick Latent Variable *z* (urn): $P_t(z)$
  - Pick frequency *f* from urn: $P(f|z)$
  - Repeat the process $\mathbf{V}_t$ times, the total energy in the *t*-th frame

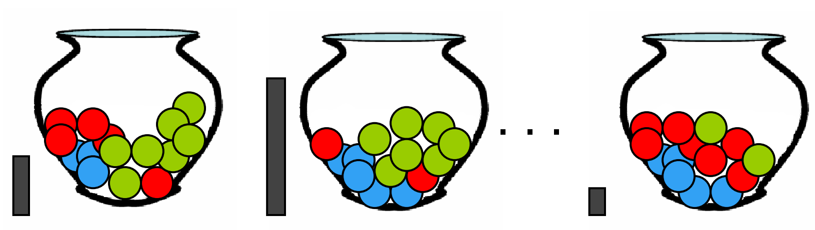
Model

Frame-specific spectral distribution

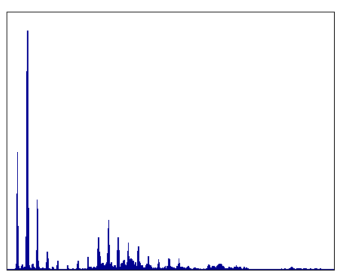
Frame-specific mixture weights

$$P_t(f) = \sum_z P(f|z)P_t(z)$$

Source-specific basis components



HISTOGRAM



f

Generative Model

- Mixture Multinomial
- Procedure
  - Pick Latent Variable *z* (urn): $P_t(z)$
  - Pick frequency *f* from urn: $P(f|z)$
  - Repeat the process $\mathbf{V}_t$ times, the total energy in the *t*-th frame

Model

Frame-specific spectral distribution

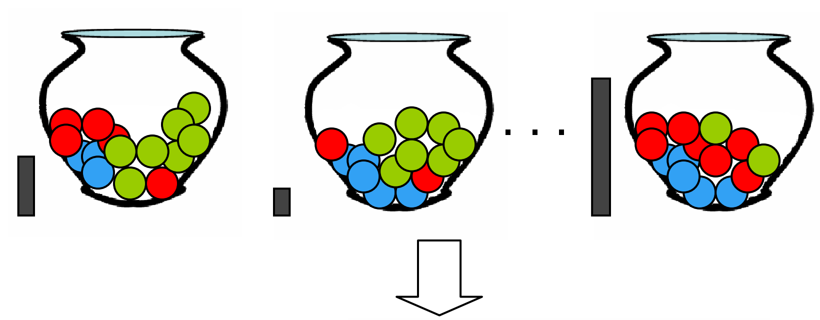
Frame-specific mixture weights

$$P_t(f) = \sum_z P(f|z) P_t(z)$$

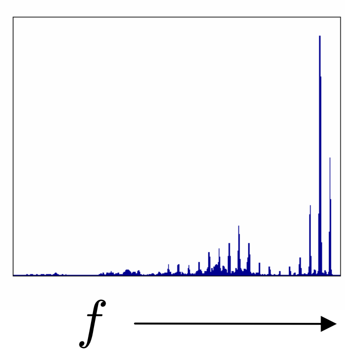
Source-specific basis components

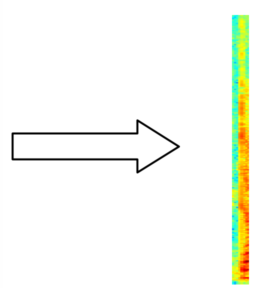
Generative Model

- Mixture Multinomial
- Procedure
  - Pick Latent Variable *z* (urn): $P_t(z)$
  - Pick frequency *f* from urn: $P(f|z)$
  - Repeat the process $\mathbf{V}_t$ times, the total energy in the *t*-th frame



HISTOGRAM





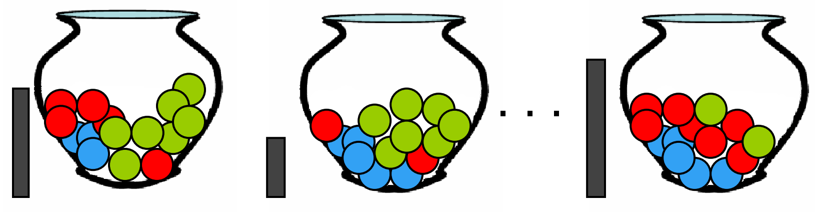
Model

Frame-specific spectral distribution

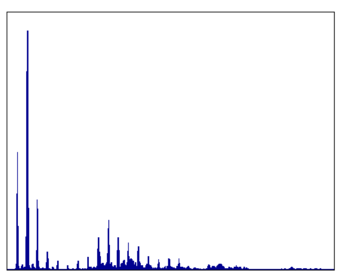
Frame-specific mixture weights

$$P_t(f) = \sum_z P(f|z)P_t(z)$$

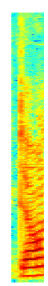
Source-specific basis components



HISTOGRAM



f

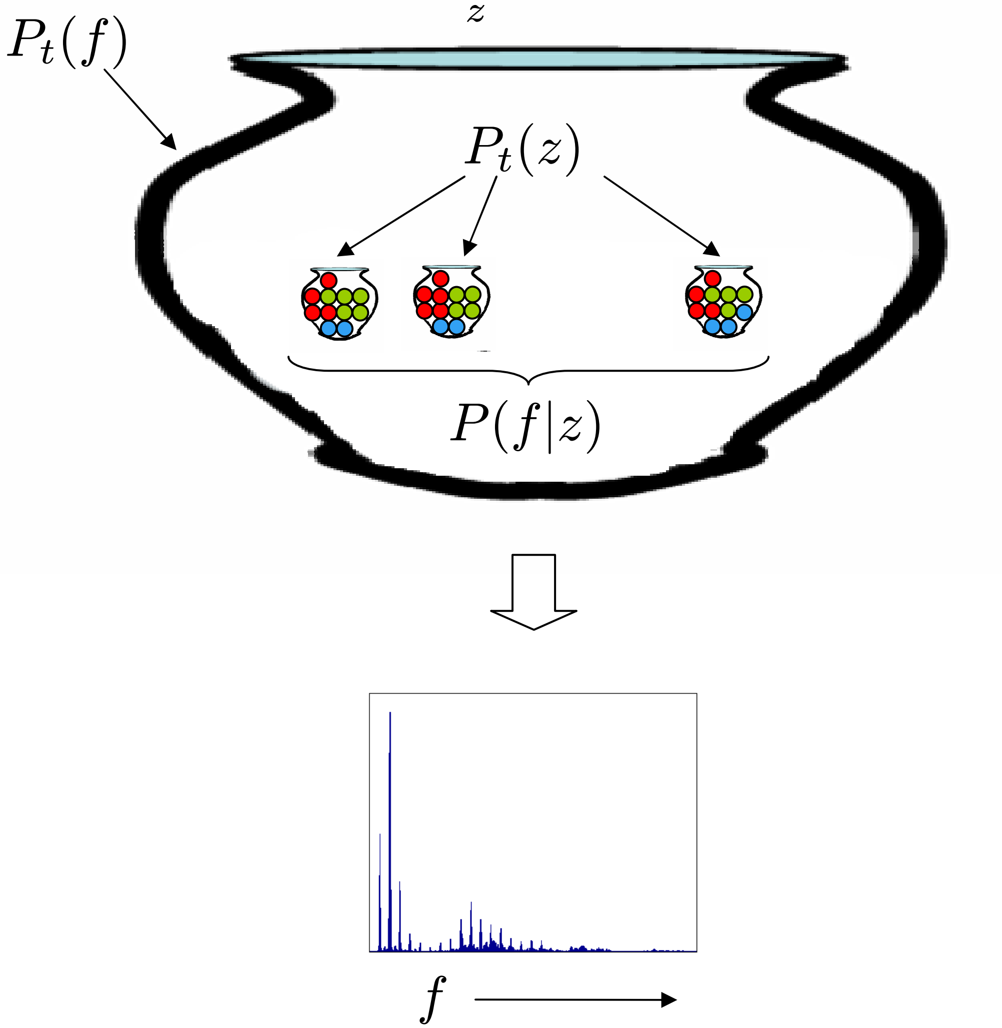


Generative Model

- Mixture Multinomial
- Procedure
  - Pick Latent Variable *z* (urn): $P_t(z)$
  - Pick frequency *f* from urn: $P(f|z)$
  - Repeat the process $\mathbf{V}_t$ times, the total energy in the *t*-th frame

Model

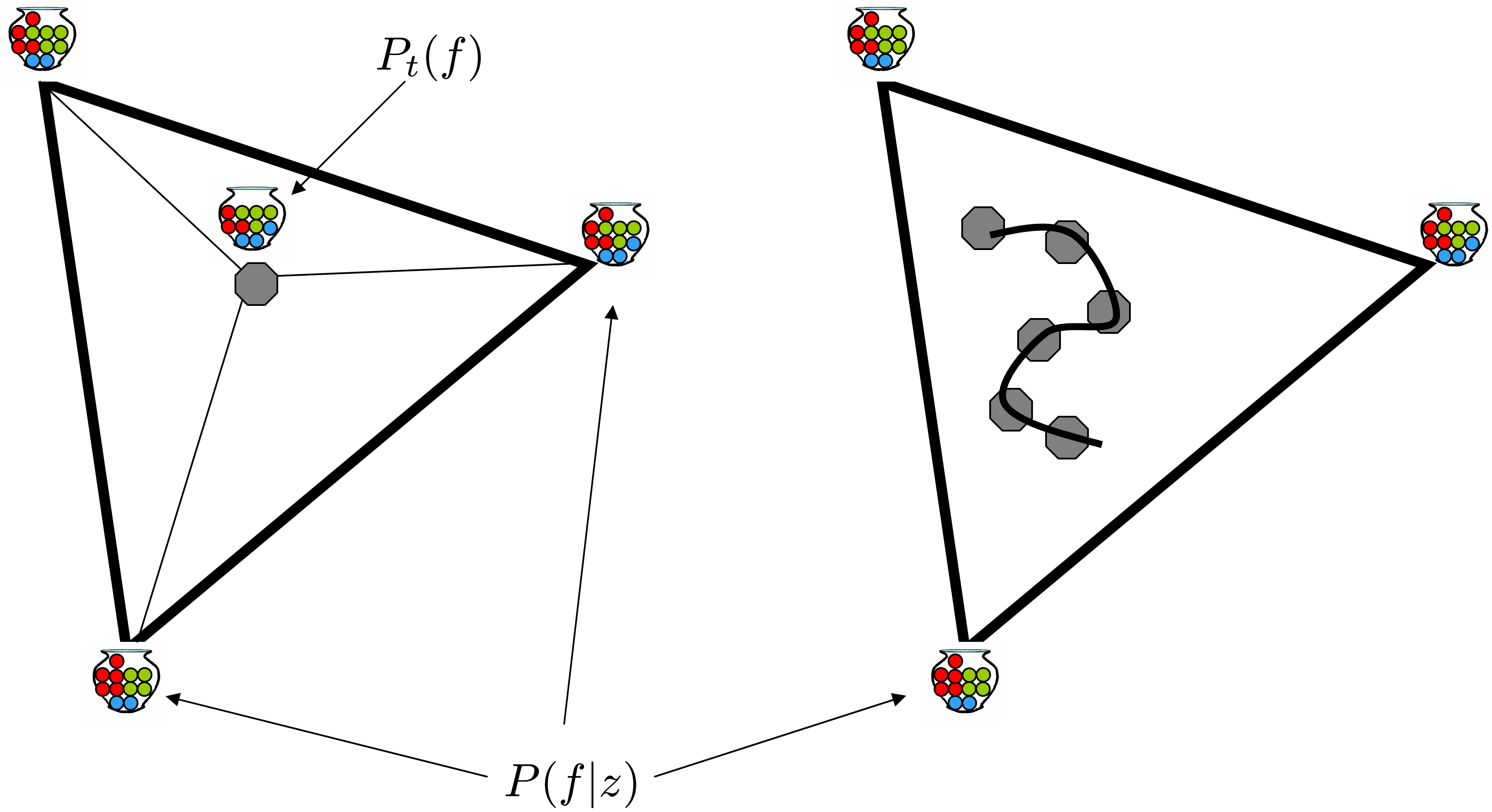
$$P_t(f) = \sum_z P(f|z)P_t(z)$$



Generative Model

- Mixture Multinomial
- Procedure
  - Pick Latent Variable *z* (urn): $P_t(z)$
  - Pick frequency *f* from urn: $P(f|z)$
  - Repeat the process $\mathbf{V}_t$ times, the total energy in the *t*-th frame

The mixture multinomial as a point in a simplex



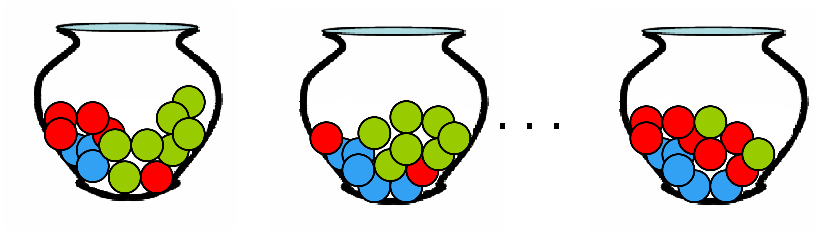
Learning the Model

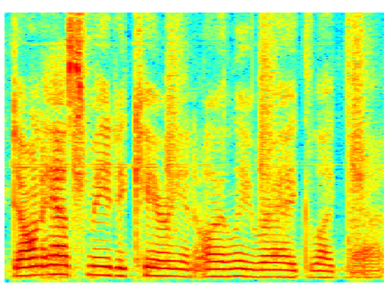
Frame-specific spectral distribution

Frame-specific mixture weights

$$P_t(f) = \sum_z P(f|z) P_t(z)$$

Source-specific basis components





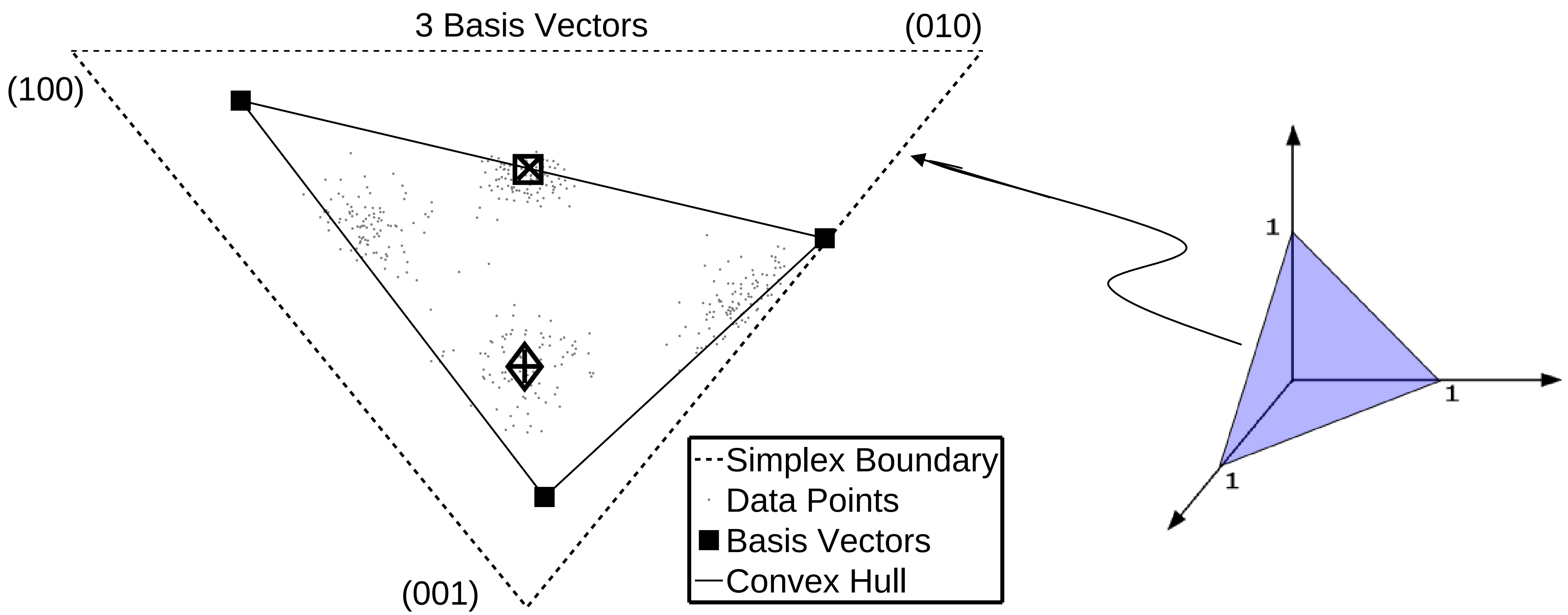
$\mathbf{V}$

Analysis

- Given the spectrogram $\mathbf{V}$, estimate the parameters

- $P(f|z)$ represent the latent structure, they underlie all the frames and hence characterize the source

Learning the Model: Geometry



- Spectral distributions and basis components are points in a simplex
- Estimation process: find corners of the convex hull that surrounds normalized spectral vectors in the simplex

Learning the Model: Parameter Estimation

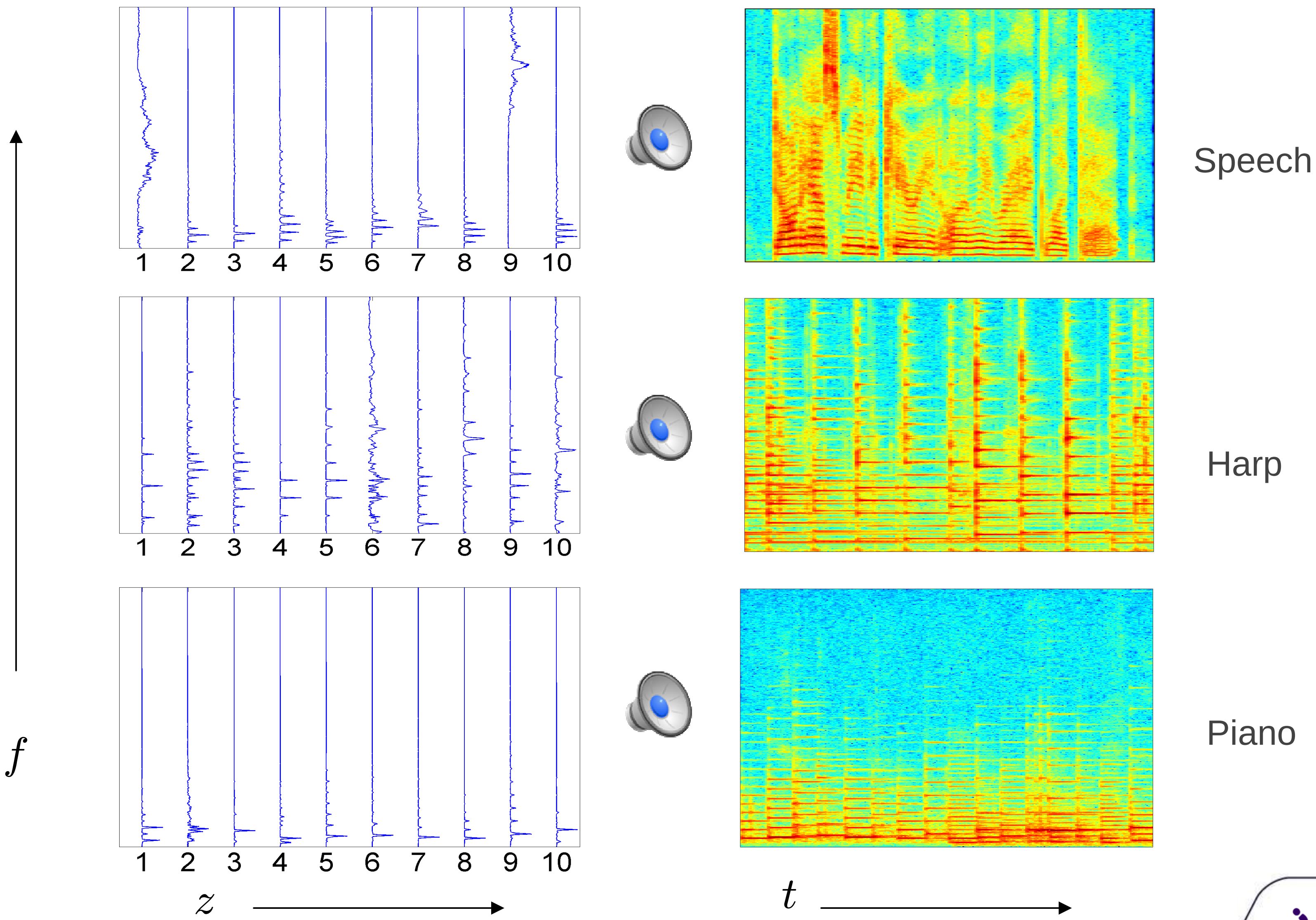
- Expectation-Maximization Algorithm

$$P_t(z|f) = \frac{P_t(z)P(f|z)}{\sum_z P_t(z)P(f|z)}$$

$$P(f|z) = \frac{\sum_t V_{ft}P_t(z|f)}{\sum_f \sum_t V_{ft}P_t(z|f)} \qquad P_t(z) = \frac{\sum_f V_{ft}P_t(z|f)}{\sum_z \sum_f V_{ft}P_t(z|f)}$$

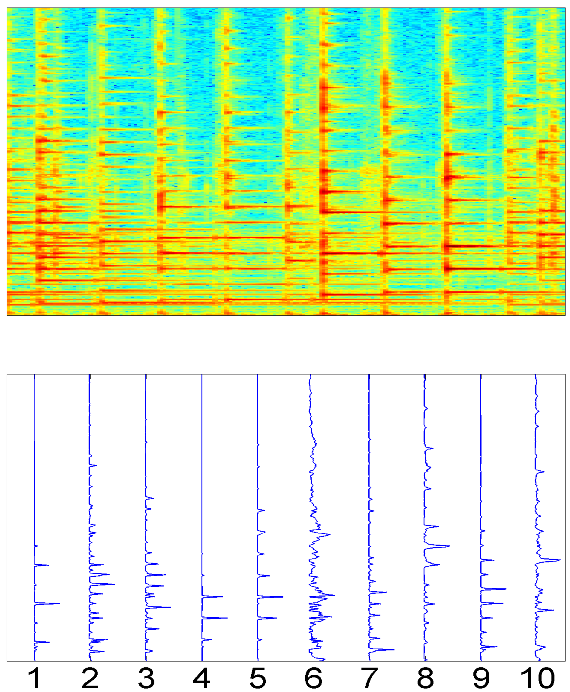
V_{ft} Entries of the training spectrogram

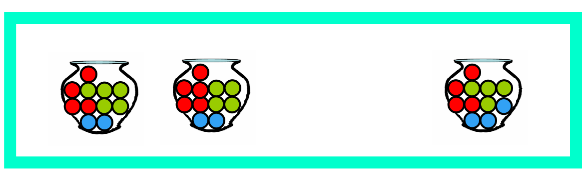
Example Bases



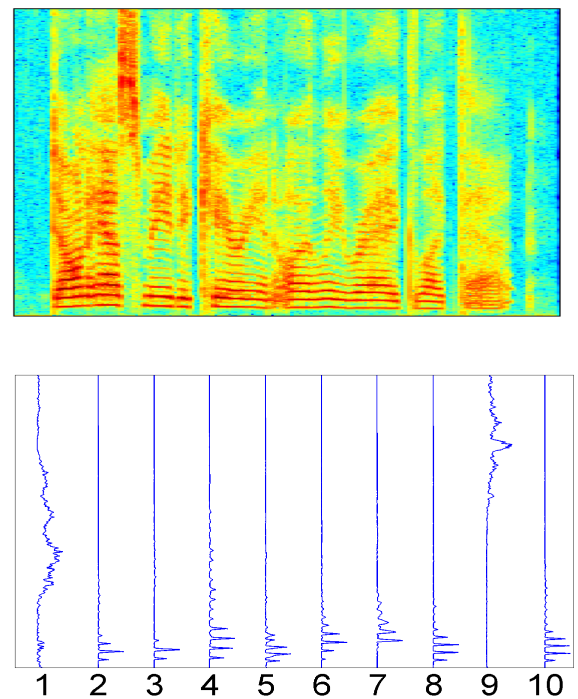
Source Separation

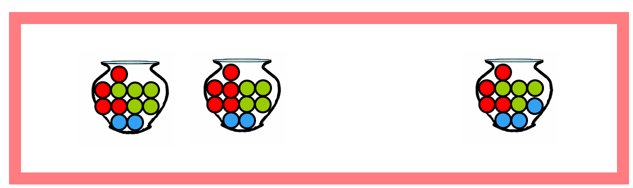
Harp





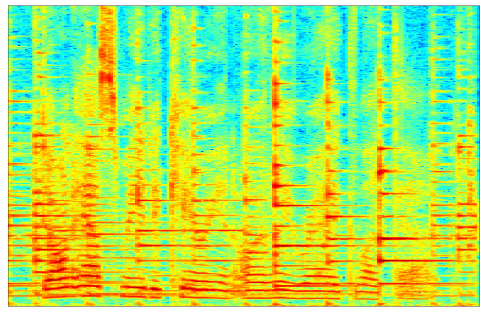
Speech



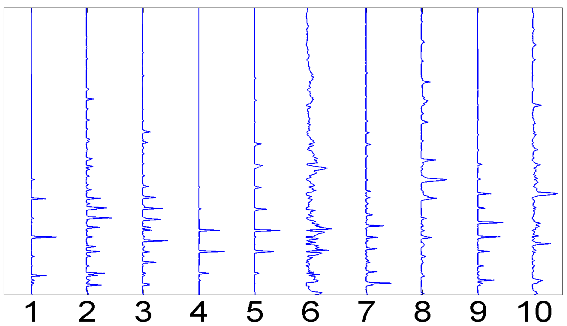


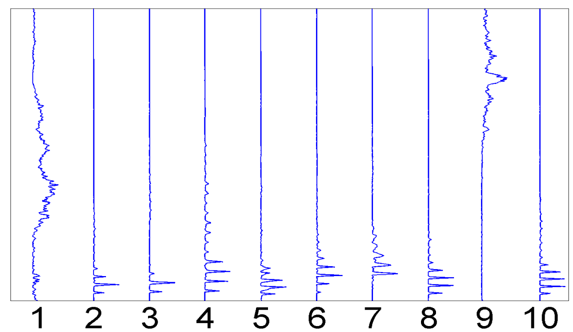
Source Separation

Mixture

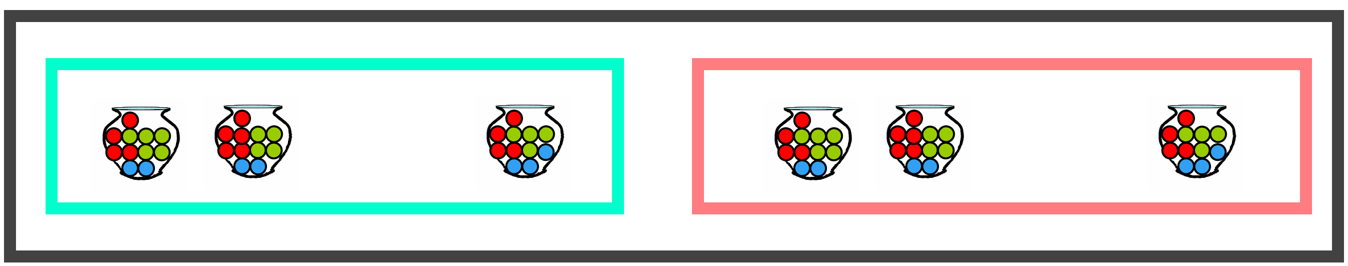


Harp Bases





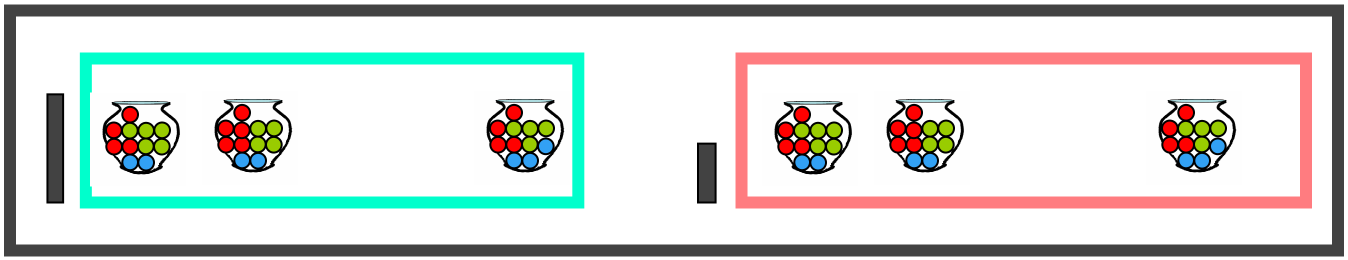
Speech Bases



Source Separation

- Mixture Spectrogram Model
  - linear combination of individual sources

$$\boxed{P_t(f)} = P_t(s_1)\boxed{P(f|s_1)} + P_t(s_2)\boxed{P(f|s_2)}$$



$$P_t(f) = \left(P_t(s_1) \sum_{z \in \{\mathbf{z}_{s_1}\}} P_t(z|s_1) P_{s_1}(f|z) \right) + \left(P_t(s_2) \sum_{z \in \{\mathbf{z}_{s_2}\}} P_t(z|s_2) P_{s_2}(f|z) \right)$$

Source Separation

- Expectation-Maximization Algorithm

$$P_t(s,z|f) = \frac{P_t(s)P_t(z|s)P_s(f|z)}{\sum_s P_t(s) \sum_{z\in\{\mathbf{z}_s\}} P_t(z|s)P_s(f|z)}$$

$$P_t(s) = \frac{\sum_f V_{ft} \sum_{z\in\{\mathbf{z}_s\}} P_t(s,z|f)}{\sum_s \sum_f V_{ft} \sum_{z\in\{\mathbf{z}_s\}} P_t(s,z|f)}$$

$$P_t(z|s) = \frac{\sum_f V_{ft} P_t(s,z|f)}{\sum_{z\in\{\mathbf{z}_s\}} \sum_f V_{ft} P_t(s,z|f)}$$

$$P_t(f,s) = P_t(s) \sum_{z\in\{\mathbf{z}_s\}} P_t(z|s)P_s(f|z)$$

V_{ft}

Entries of the mixture spectrogram

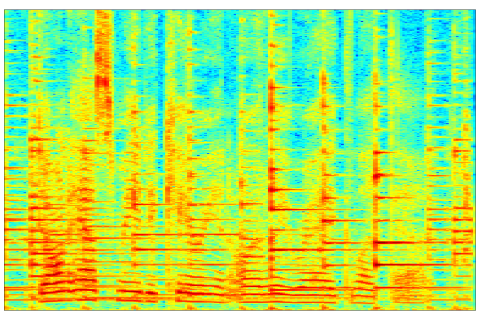
Source Separation

$$P_t(f, s)$$

$$P_t(f) = P_t(s_1)P(f|s_1) + P_t(s_2)P(f|s_2)$$

$$P_t(f) = \left(P_t(s_1) \sum_{z \in \{\mathbf{z}_{s_1}\}} P_t(z|s_1)P_{s_1}(f|z) \right) + \left(P_t(s_2) \sum_{z \in \{\mathbf{z}_{s_2}\}} P_t(z|s_2)P_{s_2}(f|z) \right)$$

V_{ft} Mixture Spectrogram



$$\hat{V}_{ft}(s) = \frac{P_t(f, s)}{P_t(f)} V_{ft}$$

Source Separation

Mixture		
Reconstructions		
SNR Improvements	5.25 dB	5.30 dB

$$SNR = g(\mathbf{R}) - g(\mathbf{V})$$

$$g(\mathbf{X}) = 10\log_{10}\left(\frac{\sum_{f,t} O_{ft}^2}{\sum_{f,t} |O_{ft}e^{j\Phi_{ft}} - X_{ft}e^{j\Phi_{ft}}|^2}\right)$$

- $\mathbf{V}$ Magnitude of the mixture
- $\mathbf{\Phi}$ Phase of the mixture
- $\mathbf{O}$ Mag. of the original signal
- $\mathbf{R}$ Mag. of the reconstruction

Source Separation: Semi supervised

- Possible even if only one source is "known"
  - Bases of other source estimated during separation

Song FG BG

- "Raise My Rent" by David Gilmour
- Background music "bases" learned from 5 seconds of music-only clips in the song
- Lead guitar bases learned from the rest of the song

Song FG BG

- "Sunrise" by Norah Jones
- Harder – wave-file clipped
- Background music bases learned from 5 seconds of music-only segments of the song
- More eg: http://cns.bu.edu/~mvss/courses/speechseg/

= +

Denoising

Only speech known

Outline

- Introduction
- Time-Frequency Structure
- Latent Variable Decomposition: Probabilistic Framework
- Sparse Overcomplete Decomposition
  - Learning more structure than the dimensionality will allow
- Conclusions

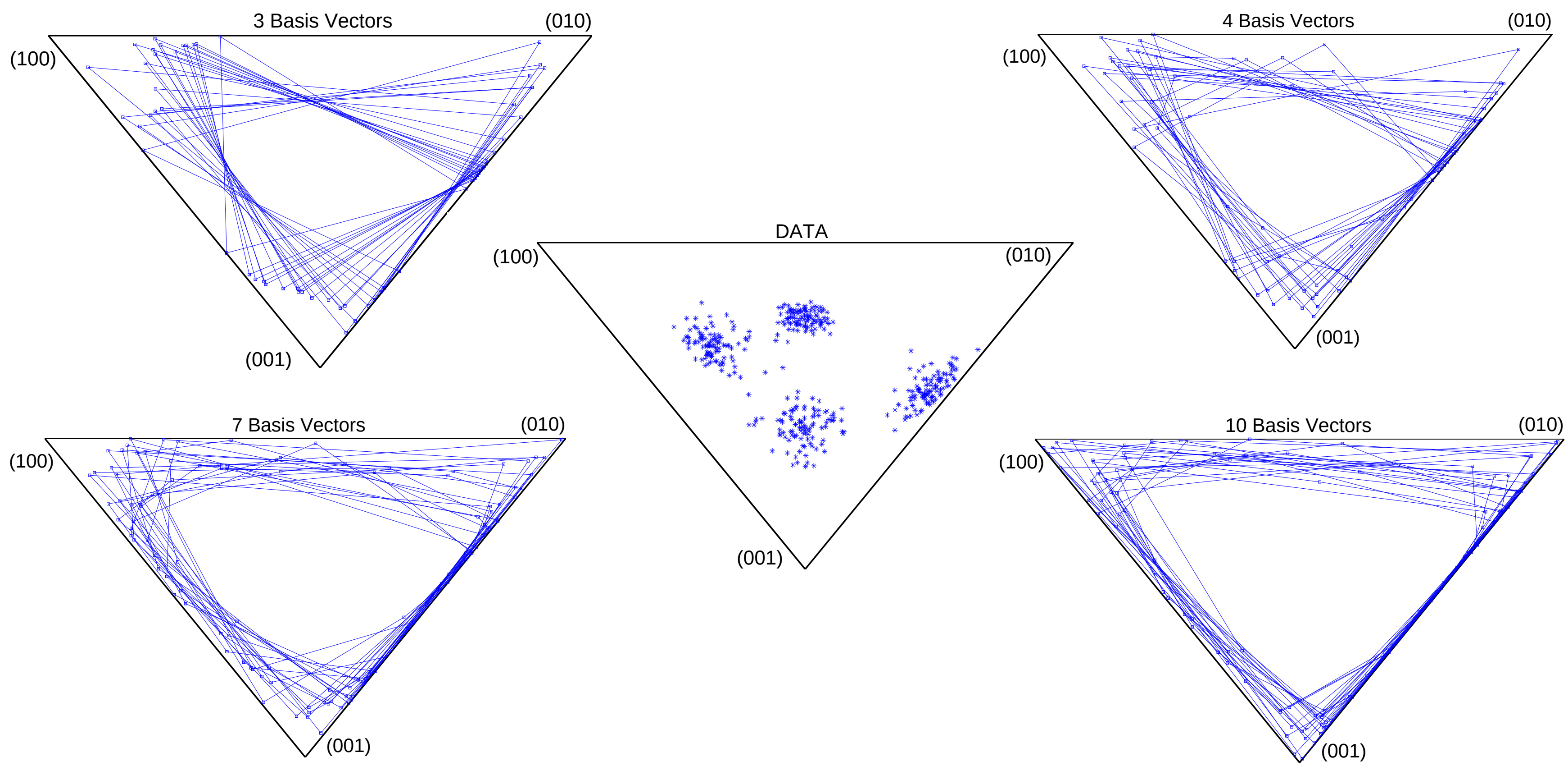
Limitation of the Framework

- Real signals exhibit complex spectral structure
  - The number of components required to model this structure could be potentially large
  - However, the latent variable framework has a limitation:

    The number of components that can be extracted is limited by the number of frequency bins in the TF representation (an arbitrary choice in the context of ground truth).

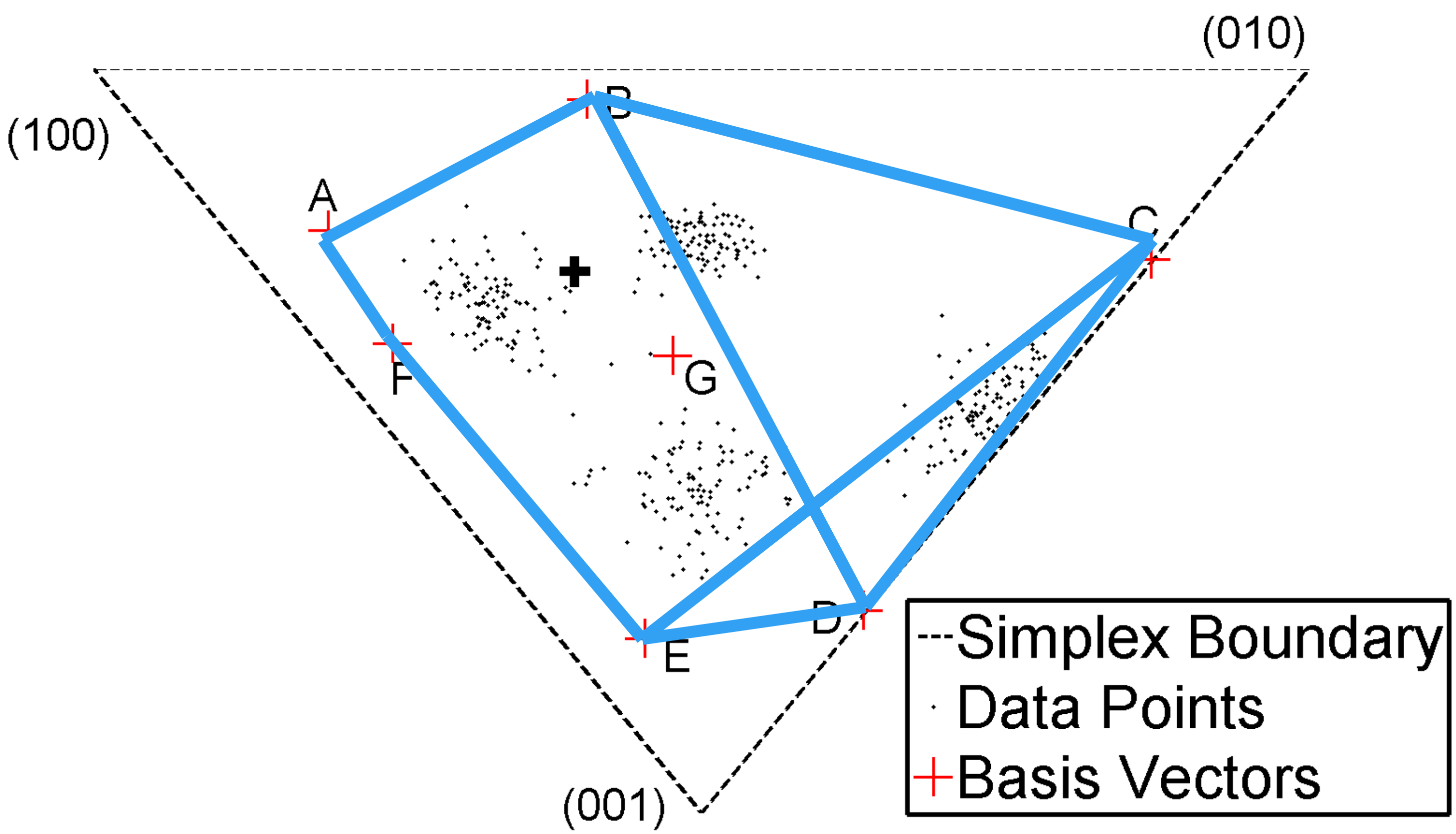
  - Extracting an *overcomplete* set of components leads to the problem of indeterminacy

Overcompleteness: Geometry



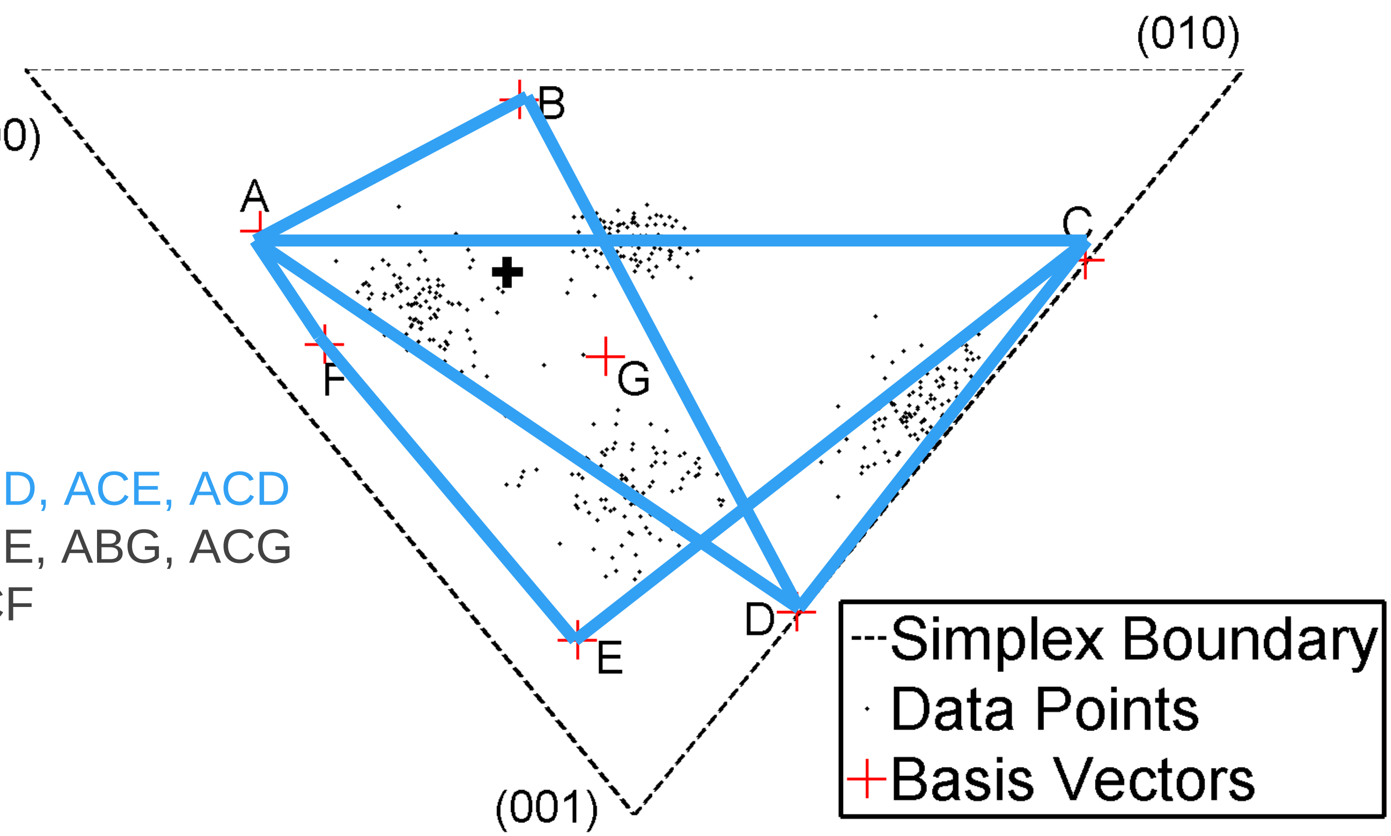
- Overcomplete case
  - As the number of bases increases, basis components migrate towards the corners of the simplex
  - Accurately represent data, but lose *data-specificity*

Indeterminacy in the Overcomplete Case



- Multiple solutions that have zero error → indeterminacy

Sparse Coding



ABD, ACE, ACD
ABE, ABG, ACG
ACF

- Restriction → use the fewest number of corners
  - At least three required for accurate representation
  - The number of possible solutions greatly reduced, but still indeterminate
  - Instead, we minimize the entropy of mixture weights

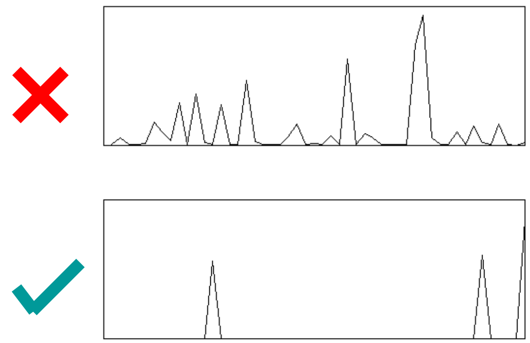
Sparsity

- Sparsity – originated as a theoretical basis for sensory coding (Kanerva, 1988; Field, 1994; Olshausen and Field, 1996)
  - Following Attneave (1954), Barlow (1959, 1961) to use information-theoretic principles to understand perception
  - Has utility in computational models and engineering methods

- How to measure sparsity?
  - fewer number of components → more sparsity
    - Number of non-zero mixture weights i.e. the L0 norm
  - L0 hard to optimize; L1 (or L2 in certain cases) used as an approximation
  - We use entropy of the mixture weights as the measure

Learning Sparse Codes: Entropic Prior

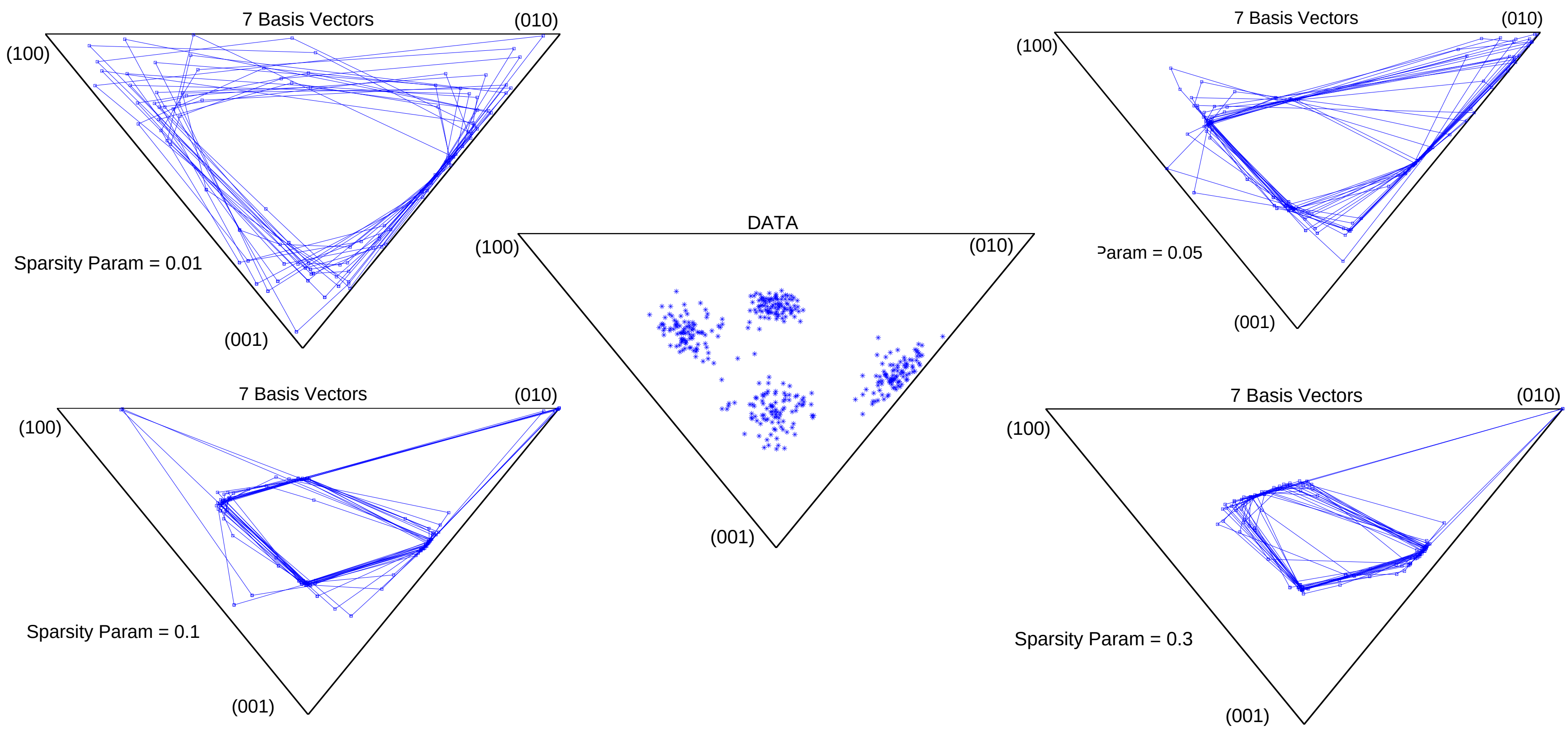
- Model -- $P_t(f) = \sum_z P(f|z) P_t(z)$
  - Estimate $P(f|z)$ such that entropy of $P_t(z)$ is minimized



$P_t(z)$

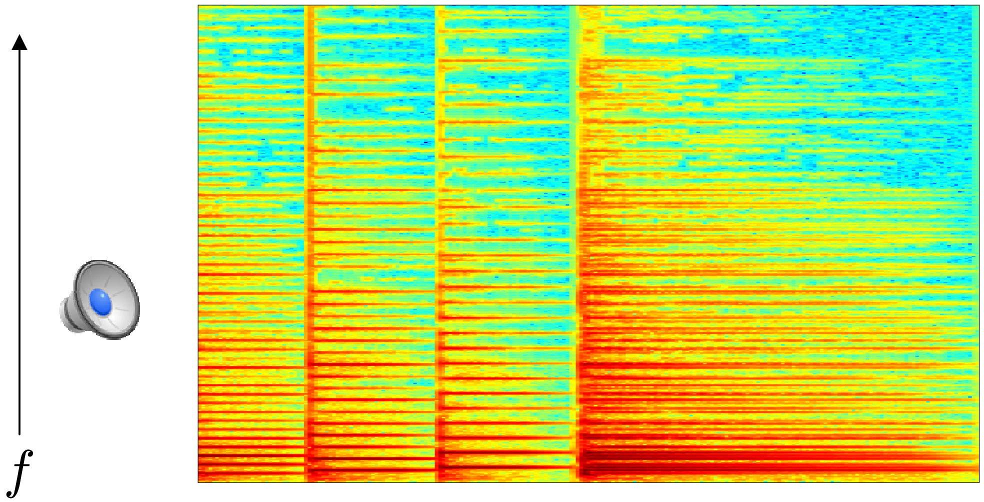
- Impose an *entropic prior* on $P_t(z)$ (Brand, 1999)
  - $P_e(\theta) \propto e^{\beta \mathcal{H}(\theta)}$ where $\mathcal{H}(\boldsymbol{\theta}) = -\sum_i \theta_i \log \theta_i$ is the entropy
  - β is the sparsity parameter that can be controlled
  - $P_t(z)$ with high entropies are penalized with low probability
  - MAP formulation solved using Lambert's W function

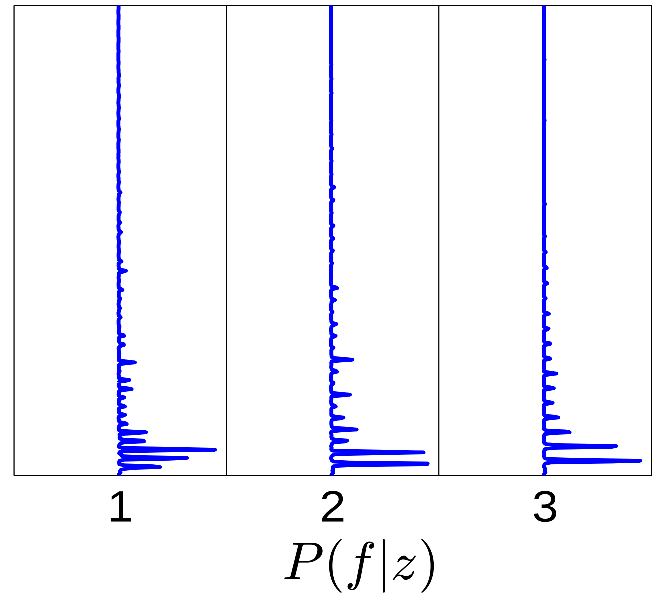
Geometry of Sparse Coding

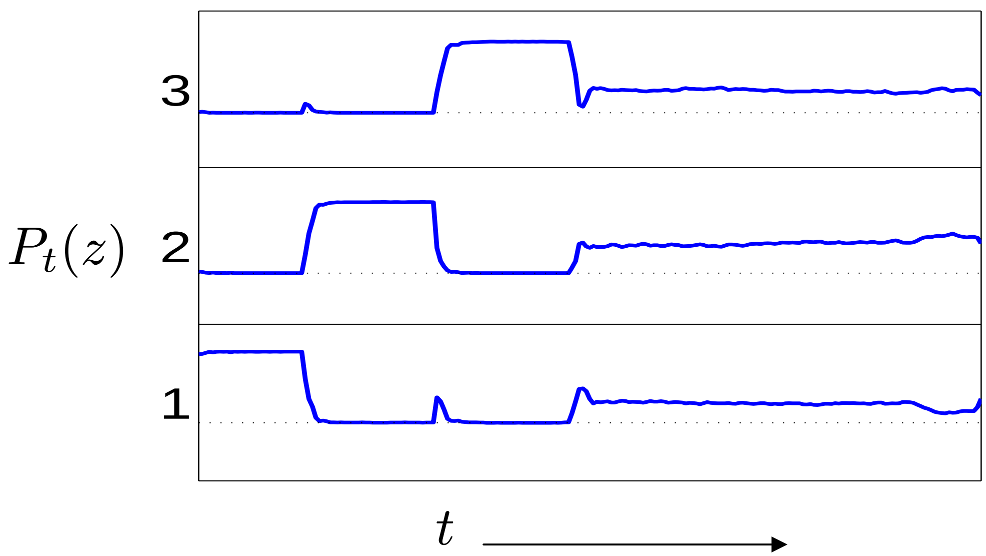


- Sparse Overcomplete case
  - Sparse mixture weights → spectral vectors must be close to a small number of corners, forcing the convex hull to be compact
  - Simplex formed by bases shrinks to fit the data

Sparse-coding: Illustration

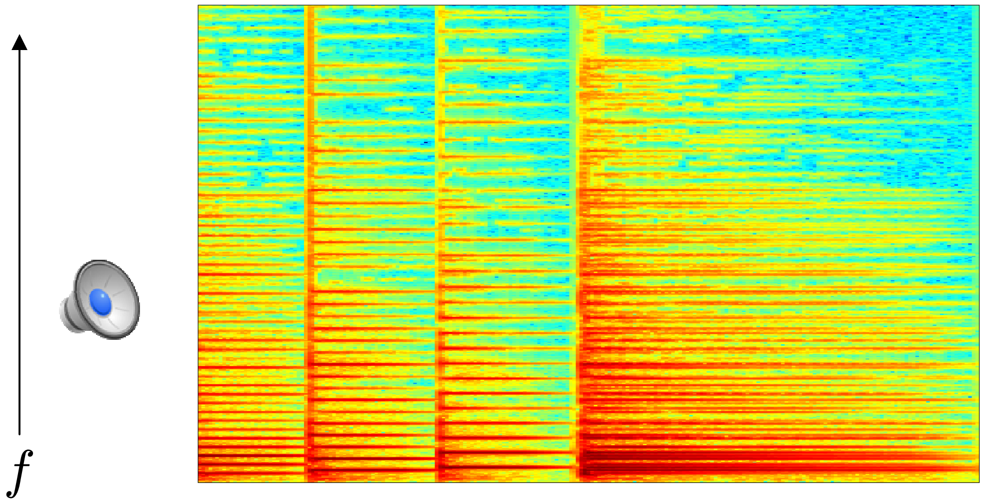


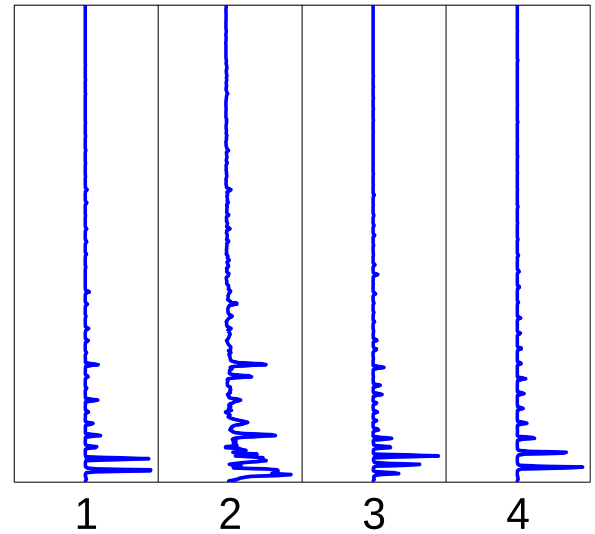


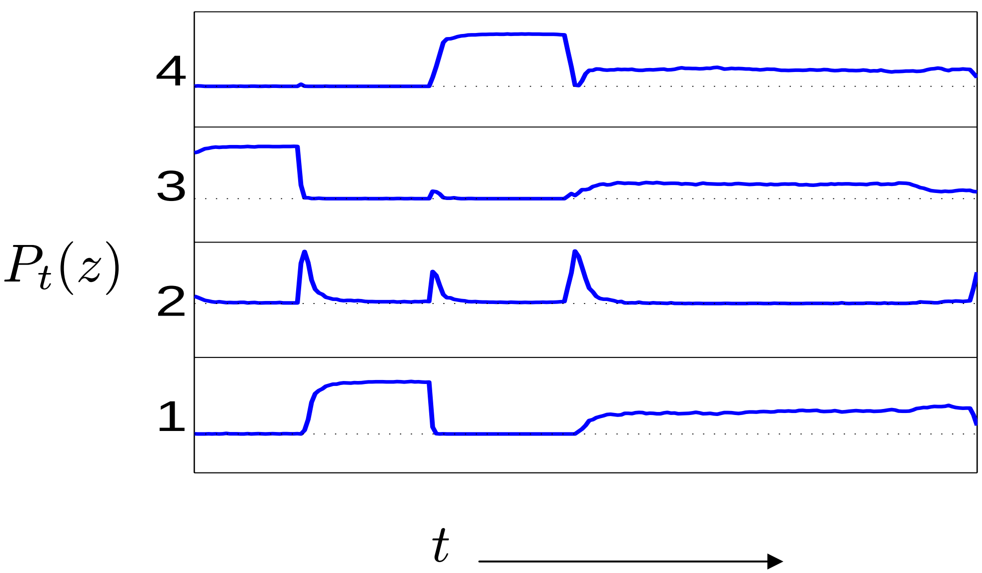


No Sparsity

Sparse-coding: Illustration

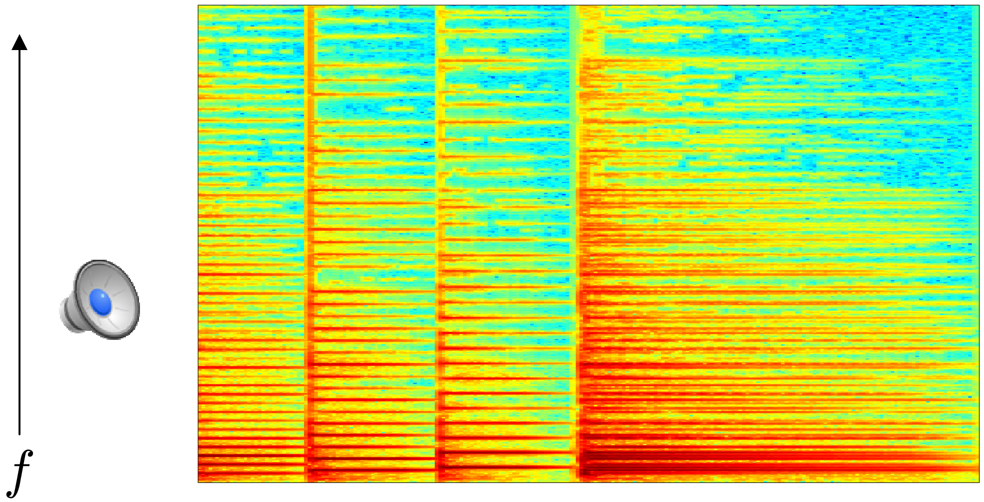


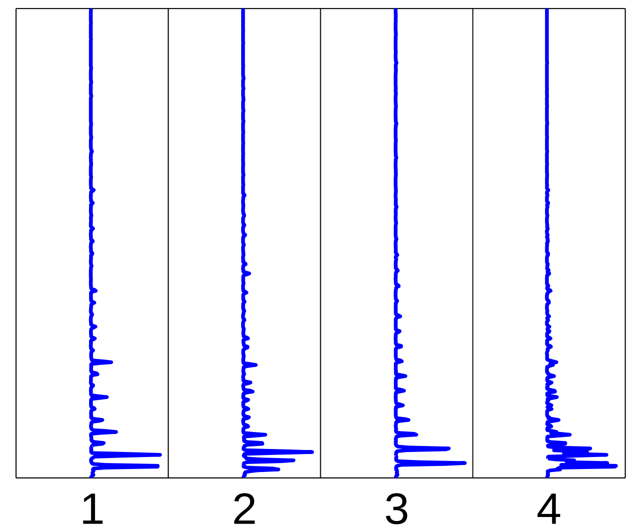




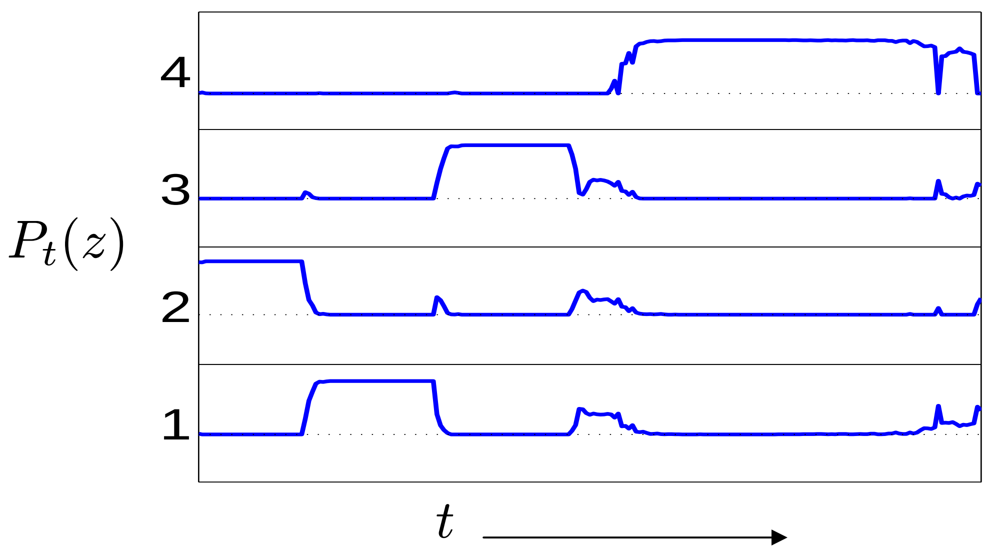
No Sparsity

Sparse-coding: Illustration





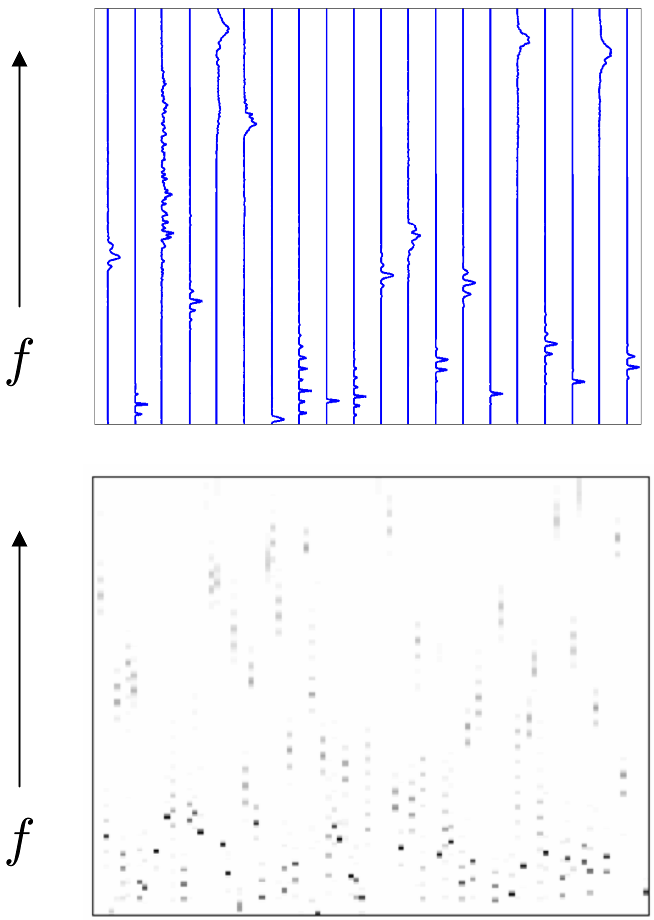
$P(f|z)$



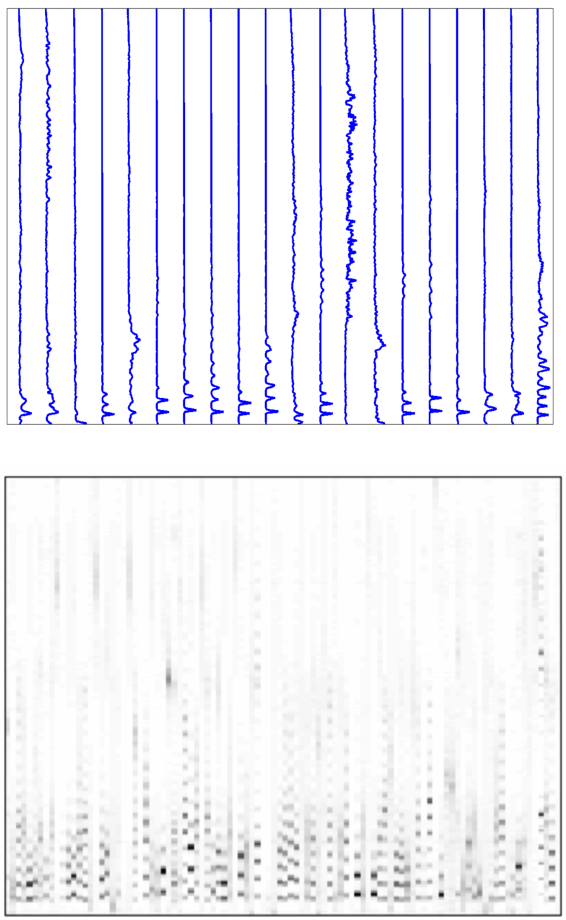
Sparse Mixture Weights

Speech Bases

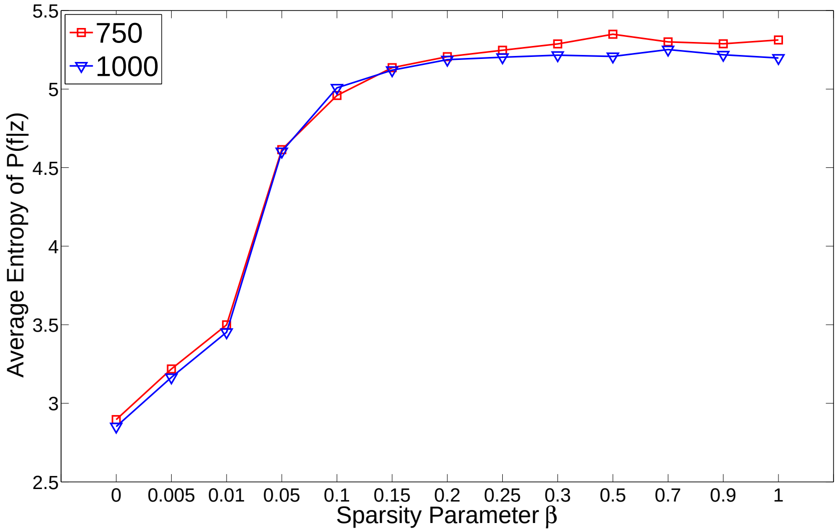
Trained **without** Sparse Mixture Weights
Compact Code

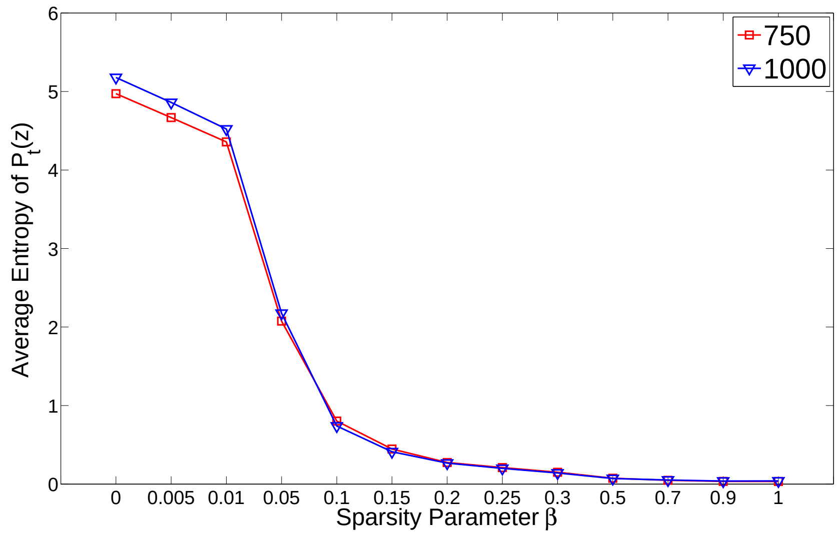


Trained **with** Sparse Mixture Weights
Sparse-Overcomplete Code



Entropy Trade-off





Sparse-coding Geometry

- Sparse mixture weights → bases which are holistic representations of the data
- Decrease in mixture-weight entropy → increase in bases components entropy, components become more "informative"
  - Empirical evidence

Source Separation: Results

Mixture 1

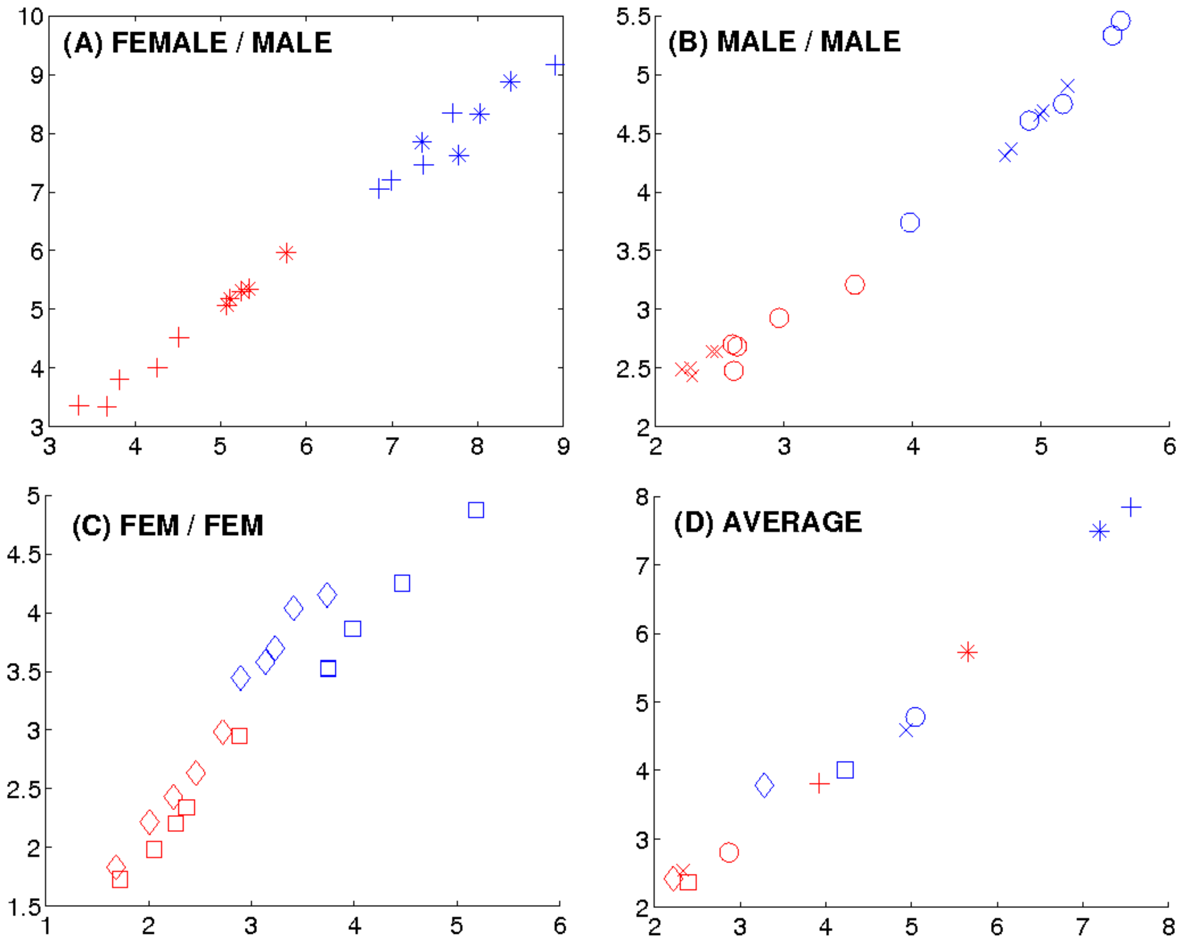
CC 3.82 dB 3.80 dB

SC 8.90 dB 9.16 dB

Mixture 2

CC 5.25 dB 5.30 dB

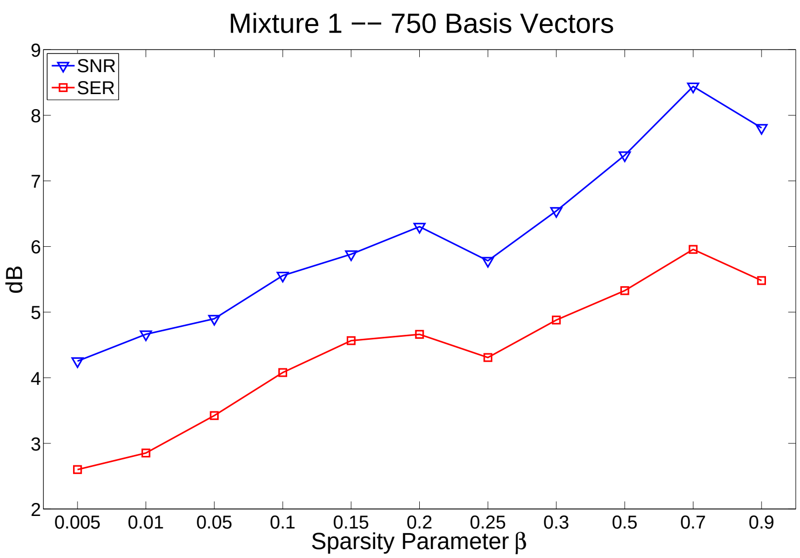
SC 8.02 dB 8.33 dB

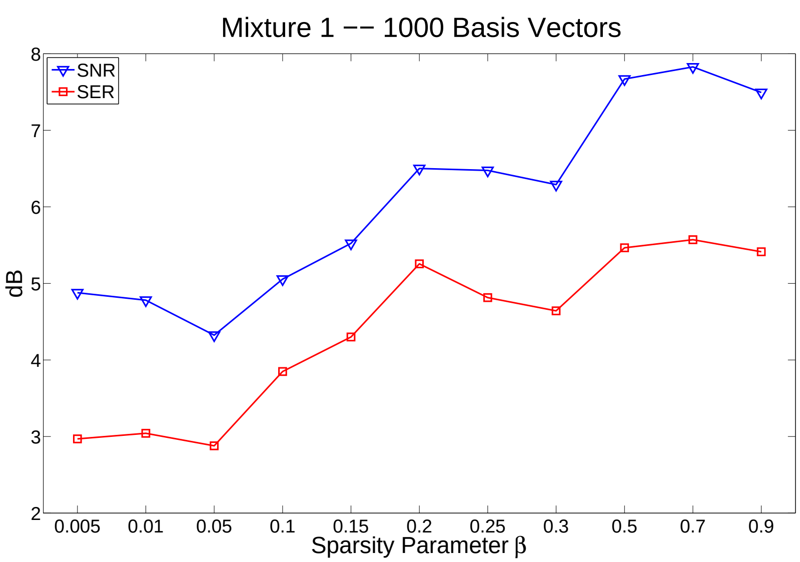


Red – CC, compact code, 100 components

Blue – SC, sparse-overcomplete code, 1000 components, $\beta = 0.7$

Source Separation: Results





Results

- Sparse-overcomplete code leads to better separation
- Separation quality increases with increasing sparsity before tapering off at high sparsity values (> 0.7)

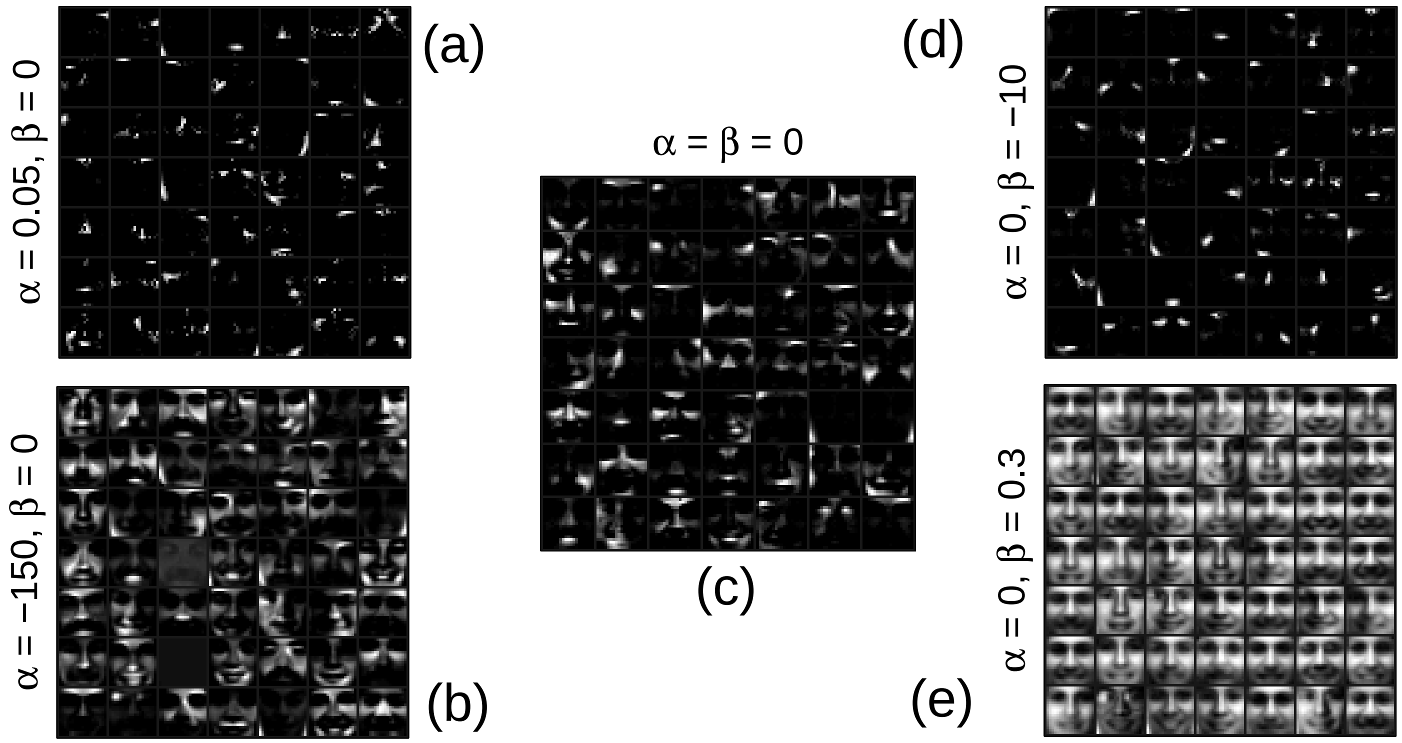
Other Applications

- Framework is general, operates on non-negative data
  - Text data (word counts), images etc.

- Examples

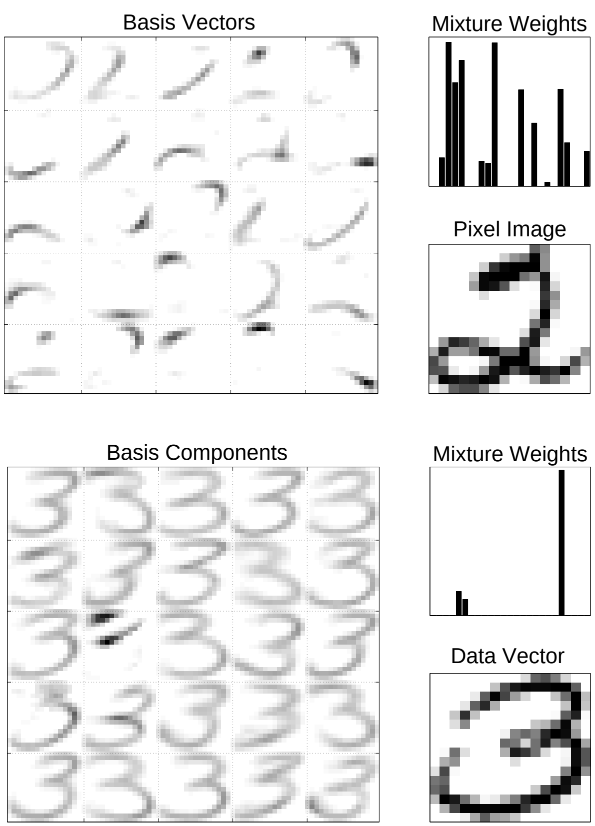
Other Applications

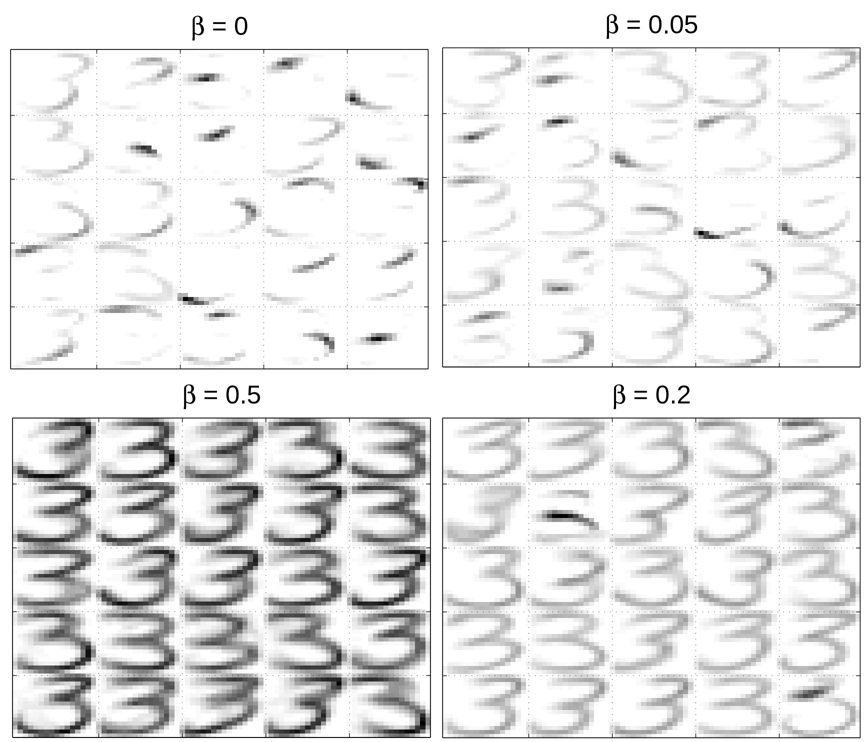
- Framework is general, operates on non-negative data
  - Text data (word counts), images etc.
- Image Examples: Feature Extraction



Other Applications

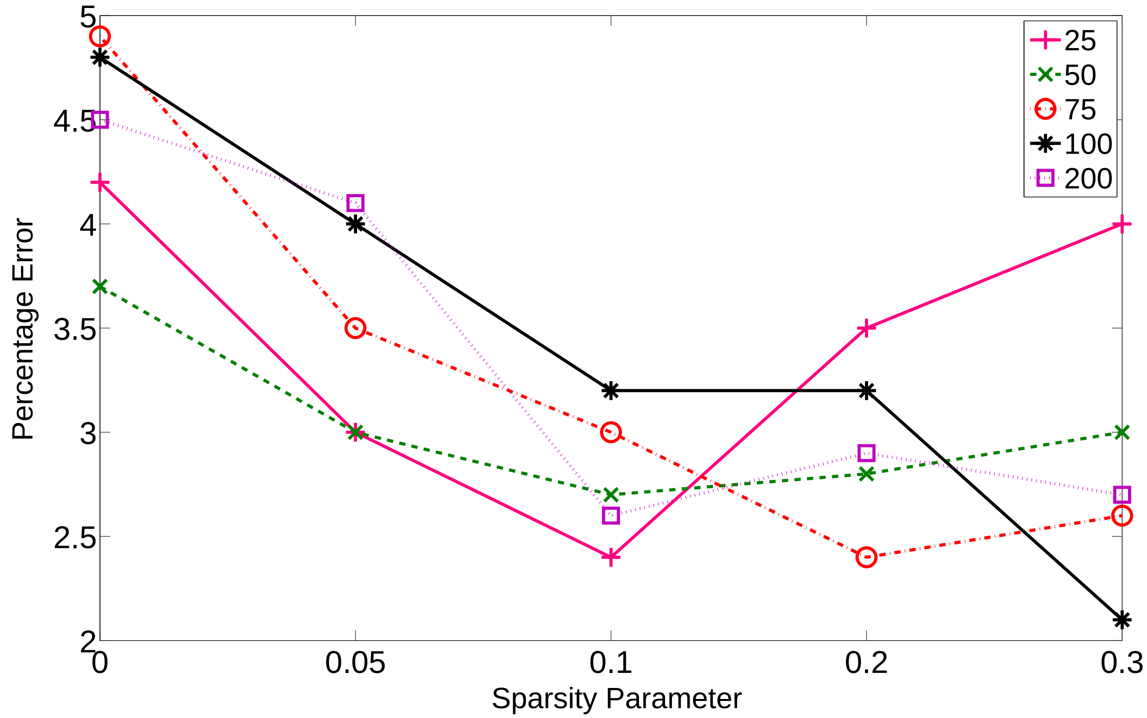
- Framework is general, operates on non-negative data
  - Text data (word counts), images etc.
- Image Examples: Hand-written digit classification





Other Applications

- Framework is general, operates on non-negative data
  - Text data (word counts), images etc.
- Image Examples: Hand-written digit classification



Outline

- Introduction
- Time-Frequency Structure
- Latent Variable Decomposition: Probabilistic Framework
- Sparse Overcomplete Decomposition
- Conclusions
  – In conclusion…

Thesis Contributions

- Modeling single-channel acoustic signals – important applications in various fields
- Provides a probabilistic framework – amenable to principled extensions and improvements
- Incorporates the idea of sparse coding in the framework
- Points to other extensions – in the form of priors
- Theoretical analysis of models and algorithms
- Applicability to other data domains

- Six refereed publications in international conferences and workshops (ICASSP, ICA, NIPS), two manuscripts under review (IEEE TPAMI, NIPS)

Future Work

- Representation
  - Other TF representations (eg. constant-Q, gammatone)
  - Multidimensional representations (correlograms, higher order spectra)
  - Utilize phase information in the representation
- Model and Theory
- Applications

Future Work

- Representation

- Model and Theory
  - Employ priors on parameters to impose known/hypothesized structure (Dirichlet, mixture Dirichlet, Logistic Normal)
  - Explicitly model time structure using HMMs/other dynamic models
  - Utilize discriminative learning paradigm
  - Extract components that form independent subspaces, could be used for unsupervised separation
  - Relation between sparse decomposition and non-negative ICA
  - Extensions/improvements to inference algorithms (eg. tempered EM)

- Applications

Future Work

- Representation
- Model and Theory
- Applications
  - Other audio applications such as music transcription, speaker recognition, audio classification, language identification etc.
  - Explore applications in data-mining, text semantic analysis, brain-imaging analysis, radiology, chemical spectral analysis etc.

Acknowledgements

- Prof. Barbara Shinn-Cunningham
- Dr. Bhiksha Raj and Dr. Paris Smaragdis
- Thesis Committee Members
- Faculty/Staff at CNS and Hearing Research Center
- Scientists/Staff at Mitsubishi Electric Research Laboratories
- Friends and well-wishers

  – Supported in part by the Air Force Office of Scientific Research (AFOSR FA9550-04-1-0260), the National Institutes of Health (NIH R01 DC05778), the National Science Foundation (NSF SBE-0354378), and the Office of Naval Research (ONR N00014-01-1-0624).

  – Arts and Sciences Dean's Fellowship, Teaching Fellowship

  – Internships at Mitsubishi Electric Research Laboratories

Additional Slides

Publications

Refereed Publications and Manuscripts Under Review

- MVS Shashanka, B Raj, P Smaragdis. "*Probabilistic Latent Variable Model for Sparse Decompositions of Non-negative Data*" submitted to IEEE Transactions on Pattern Analysis And Machine Intelligence
- MVS Shashanka, B Raj, P Sparagdis. *"Sparse Overcomplete Latent Variable Decomposition of Counts Data*" submitted to NIPS 2007
- P Smaragdis, B Raj, MVS Shashanka. *"Supervised and Semi-Supervised Separation of Sounds from Single-Channel Mixtures*," Intl. Conf. on ICA, London, Sep 2007
- MVS Shashanka, B Raj, P Smaragdis. *"Sparse Overcomplete Decomposition for Single Channel Speaker Separation,"* Intl. Conf. on Acoustics, Speech and Signal Processing, Honolulu, Apr 2007
- B Raj, R Singh, MVS Shashanka, P Smaragdis. *"Bandwidth Expansion with a Polya Urn Model,"* Intl. Conf. on Acoustics, Speech and Signal Proc., Honolulu, Apr 2007
- B Raj, P Smaragdis, MVS Shashanka, R Singh, *"Separating a Foreground Singer from Background Music,"* Intl Symposium on Frontiers of Research on Speech and Music, Mysore, India, Jan 2007
- P Smaragdis, B Raj, MVS Shashanka. *"A Probabilistic Latent Variable Model for Acoustic Modeling ,"* Workshop on Advances in Models for Acoustic Processing, NIPS 2006
- B Raj, MVS Shashanka, P Smaragdis. *"Latent Dirichlet Decomposition for Single Channel Speaker Separation,"* Intl. Conf. on Acoustics, Speech and Signal Processing, Paris, May 2006