SoundSense: Scalable Sound Sensing for People-Centric Applications on Mobile Phones

Hong Lu, Wei Pan, Nicholas D. Lane, Tanzeem Choudhury, Andrew T. Campbell
Dept. of Computer Science, Dartmouth College

Mobile Phones are the perfect platform for large scale sensor networks

Sounds captured by the microphone allow us to sense people and the environment.

SoundSense

Hong Lu, Wei Pan, Nicholas D. Lane, Tanzeem Choudhury, Andrew T. Campbell

design

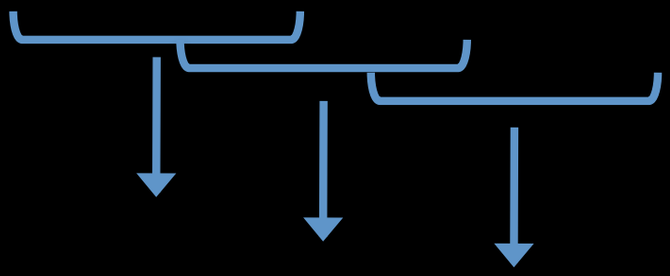
Design Issues

- Scalable Sound Classification
- Privacy
- Resource limitations
- Phone context

SoundSense Pipeline

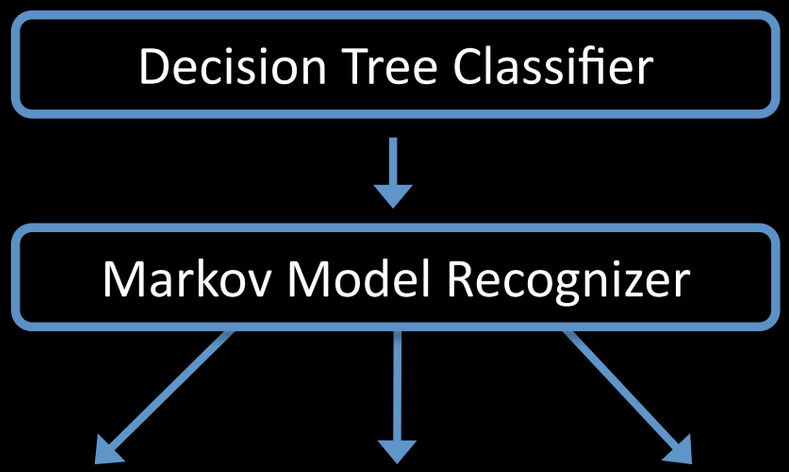
Sound Waveform

Framing



Feature Extraction

Coarse Category Classification



Intra-Category Classification

SoundSense Pipeline

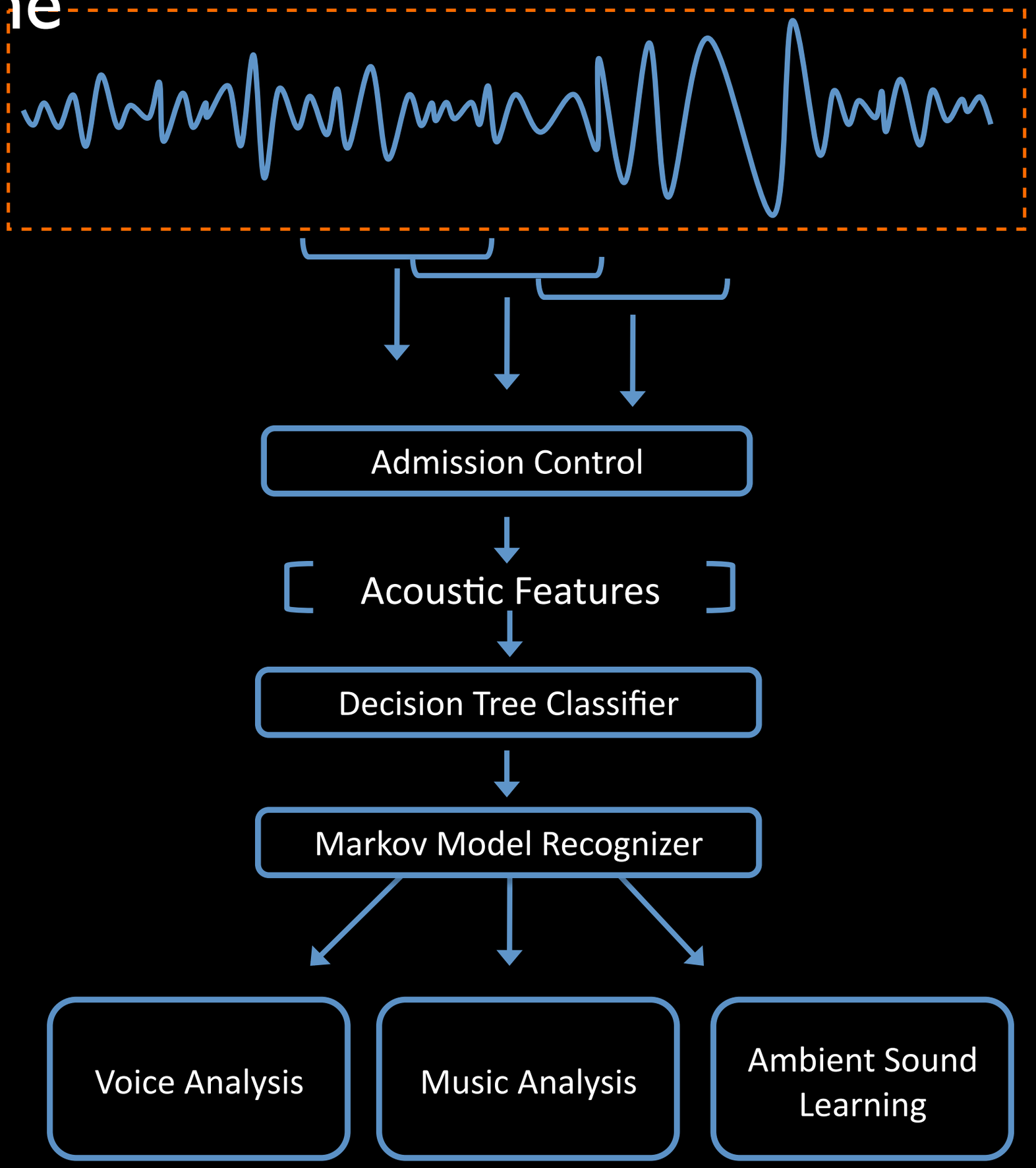
Sound Waveform

Framing

Feature Extraction

Coarse Category Classification

Intra-Category Classification



SoundSense Pipeline

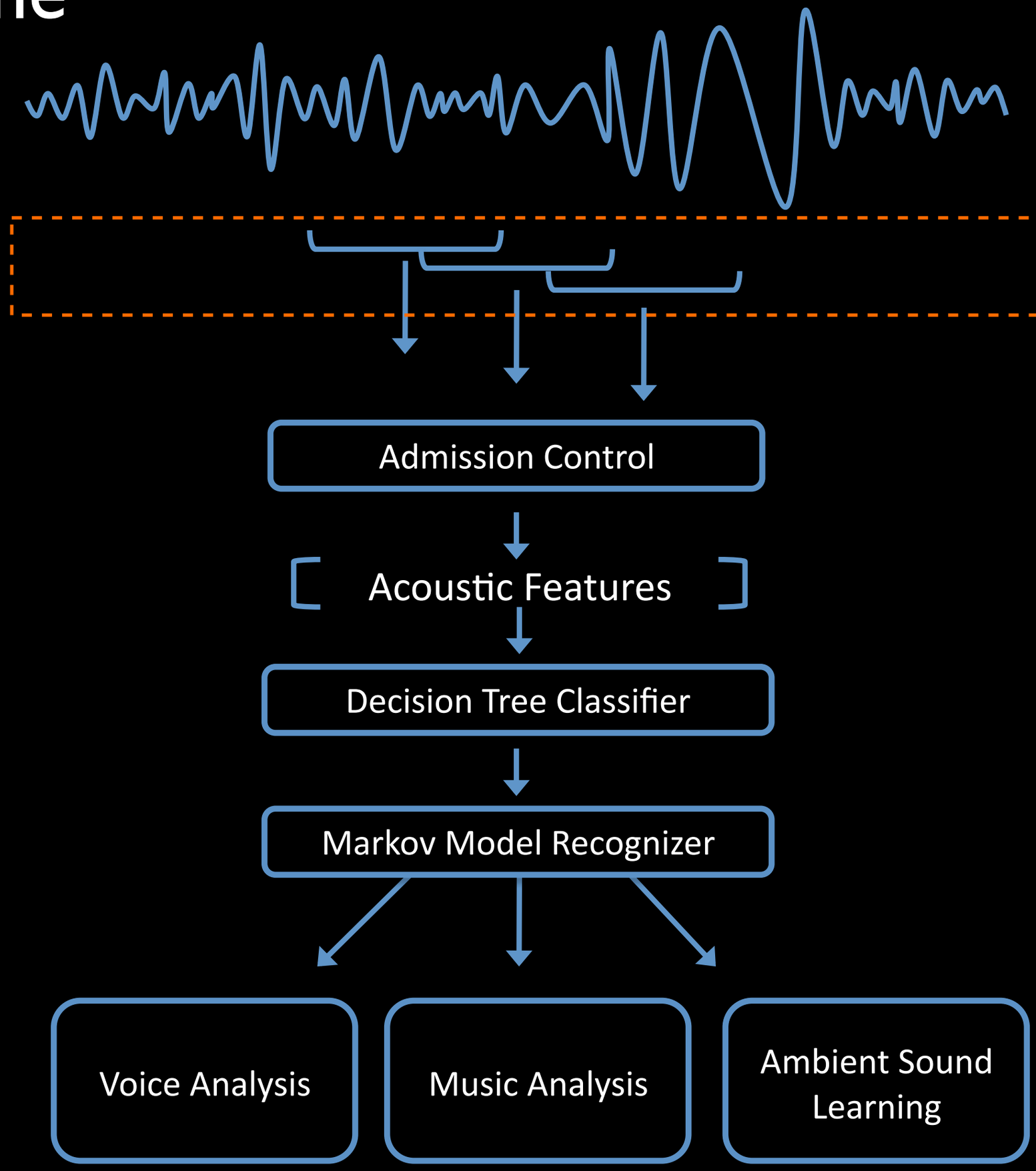
Sound Waveform

Framing

Feature Extraction

Coarse Category Classification

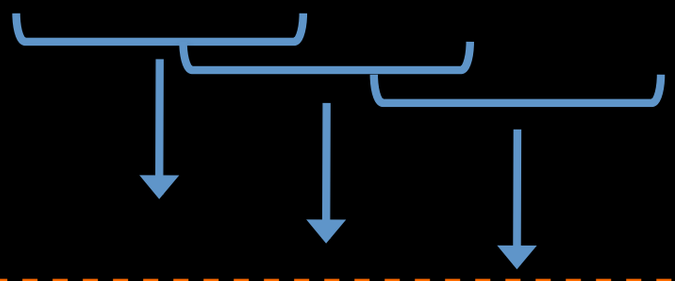
Intra-Category Classification



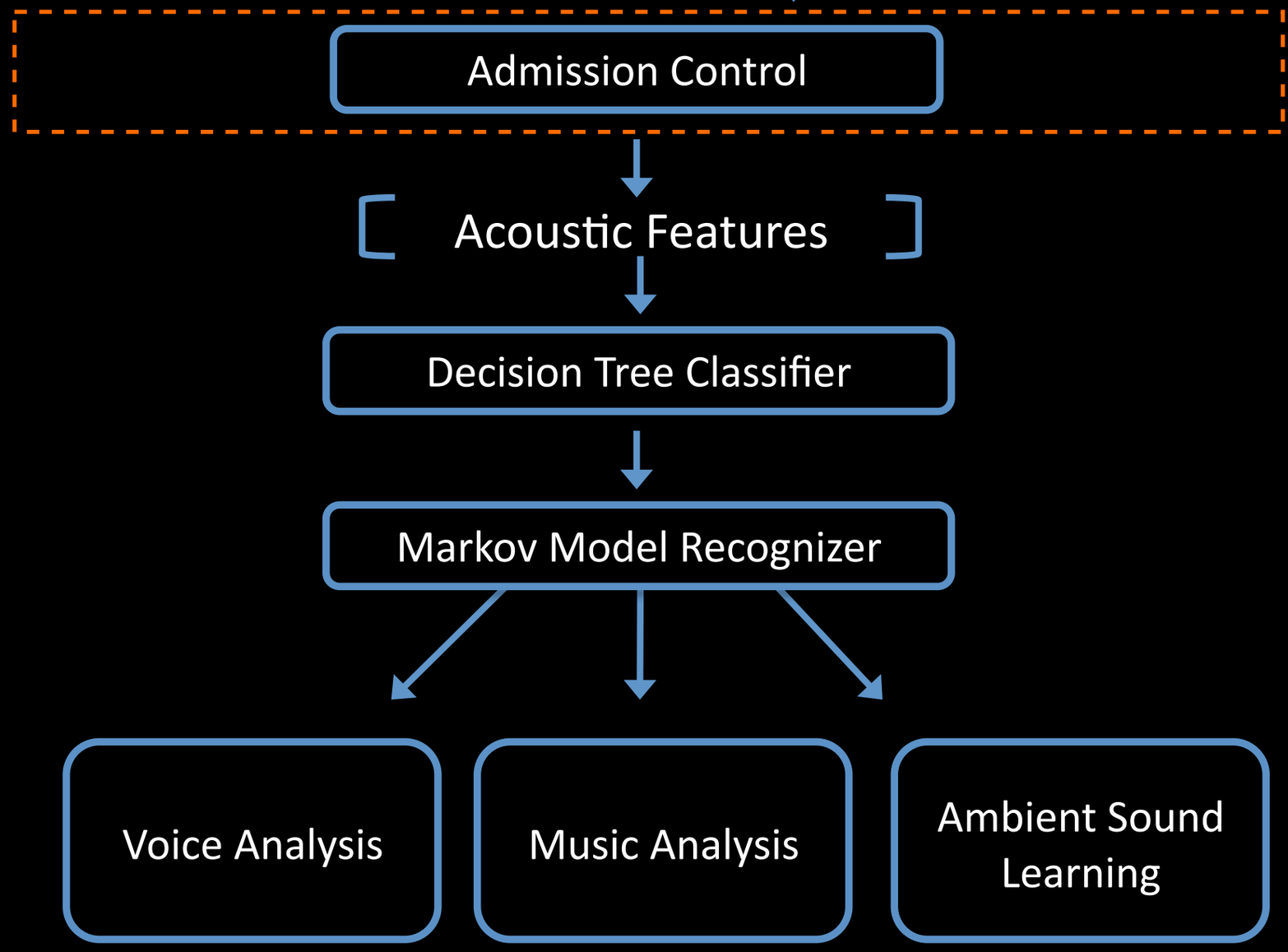
SoundSense Pipeline

Sound Waveform

Framing



Feature Extraction



Coarse Category Classification

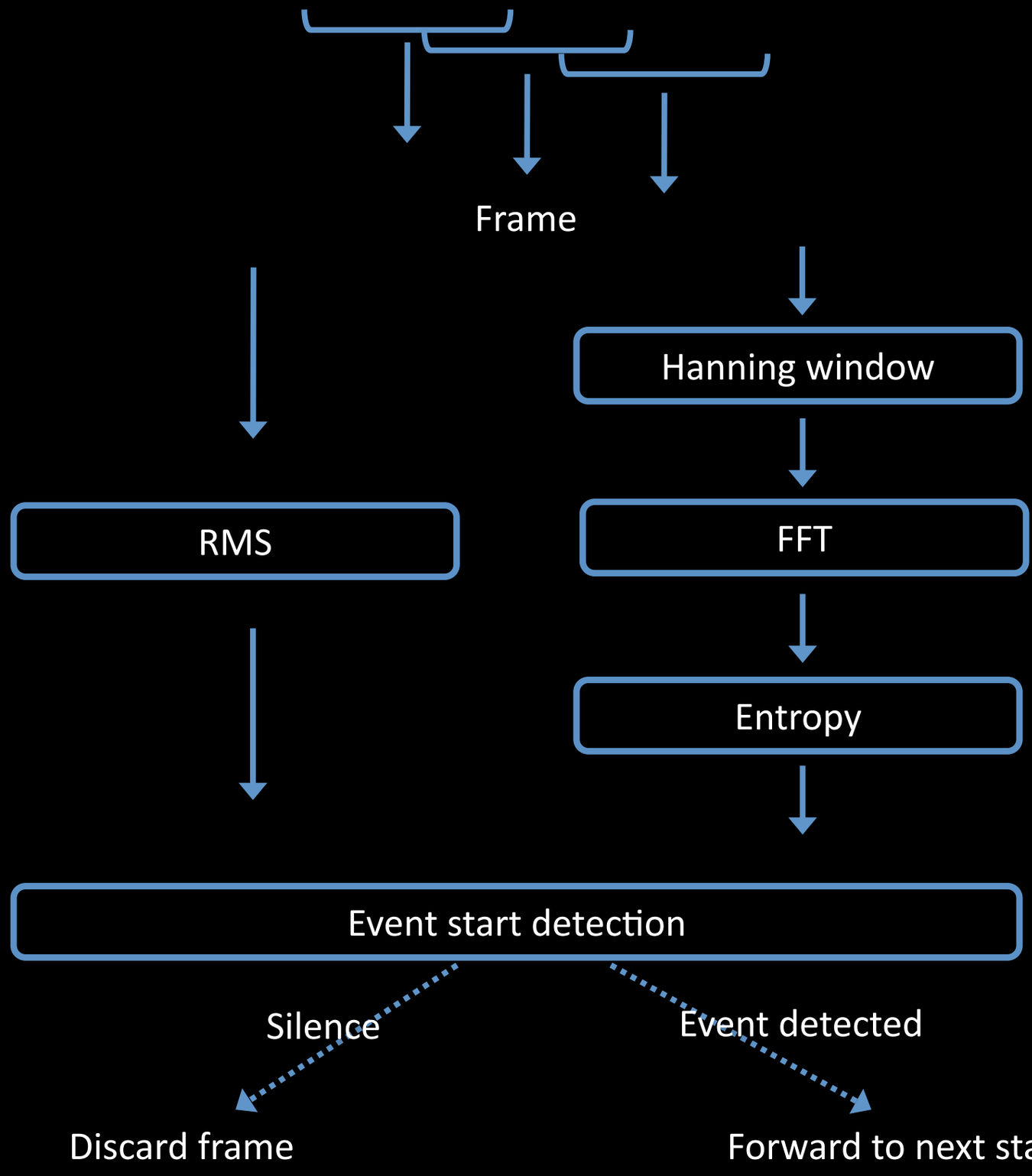
Intra-Category Classification

Admission control lowers the resources consumed during continuous sensing

- Filter out silence and non-informative frames

- Criteria:

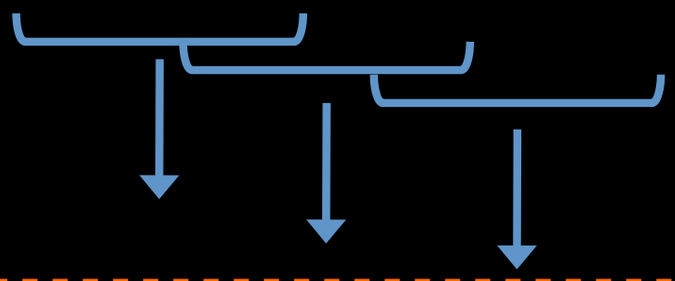
- Energy & Spectral Entropy



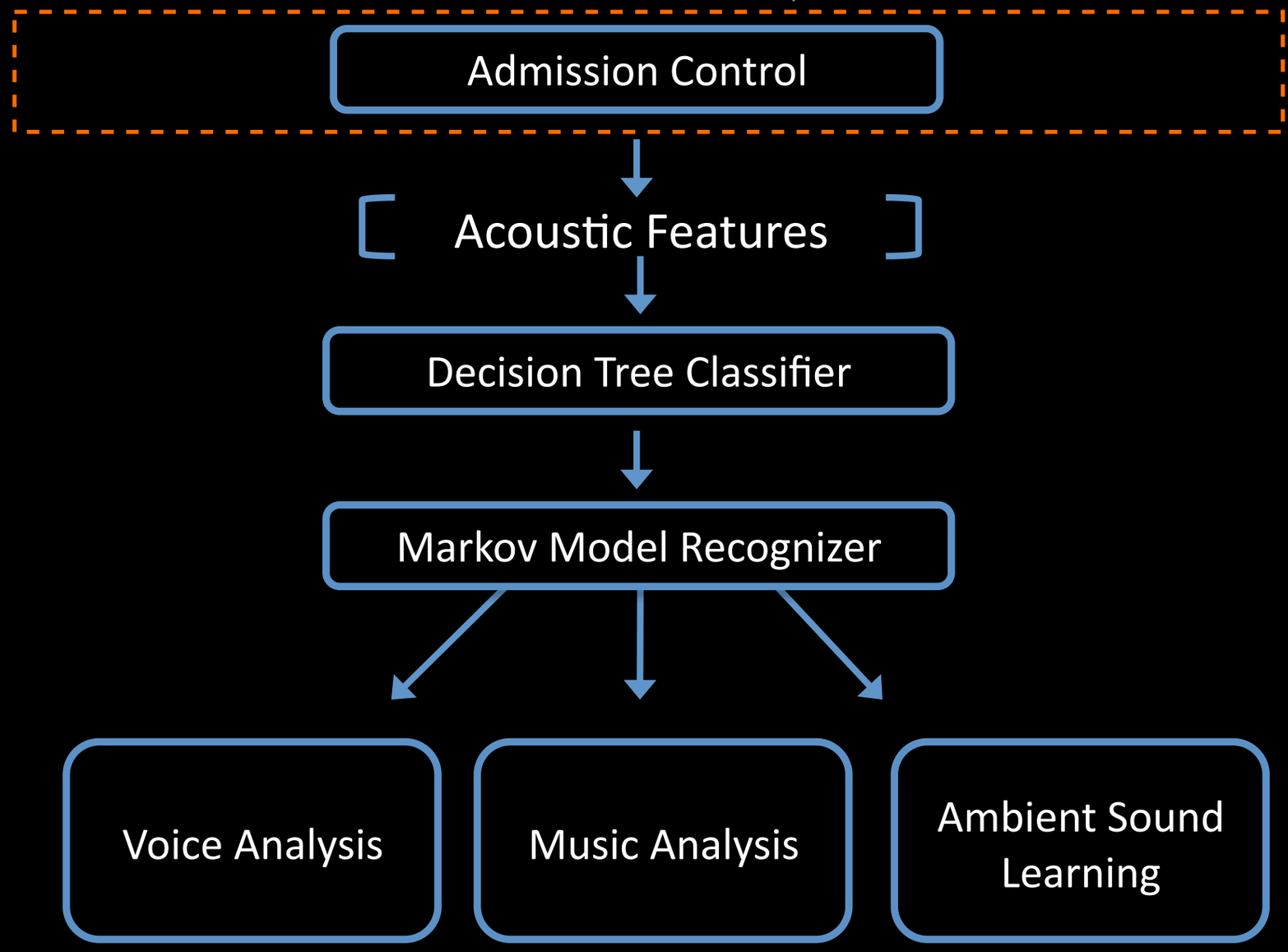
SoundSense Pipeline

Sound Waveform

Framing



Feature Extraction



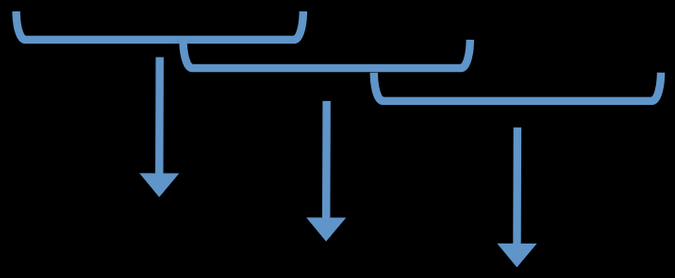
Coarse Category Classification

Intra-Category Classification

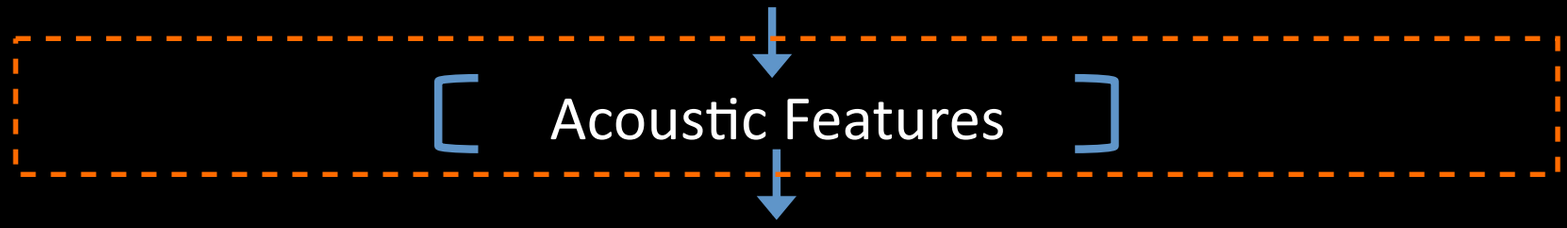
SoundSense Pipeline

Sound Waveform

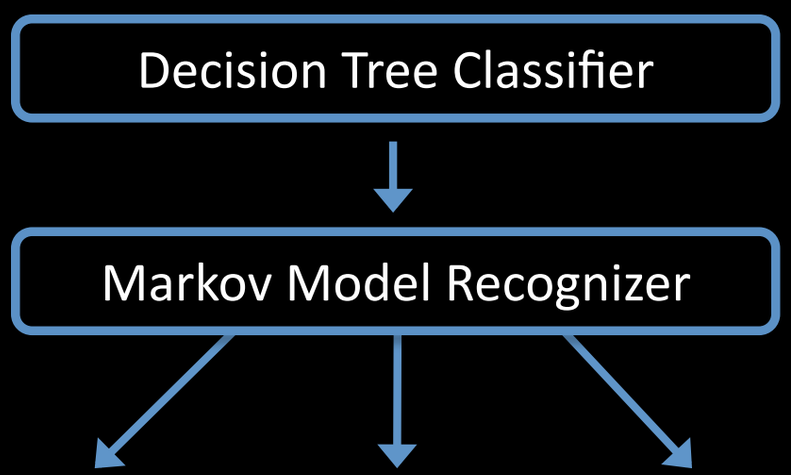
Framing



Feature Extraction

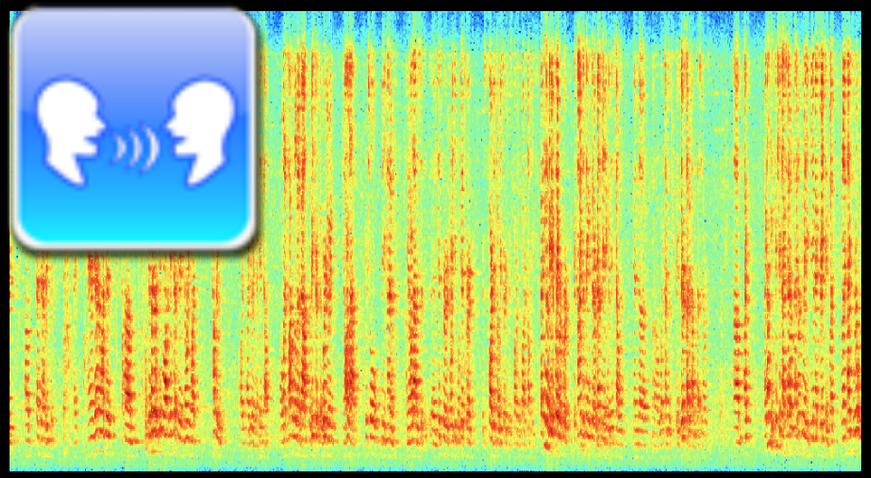


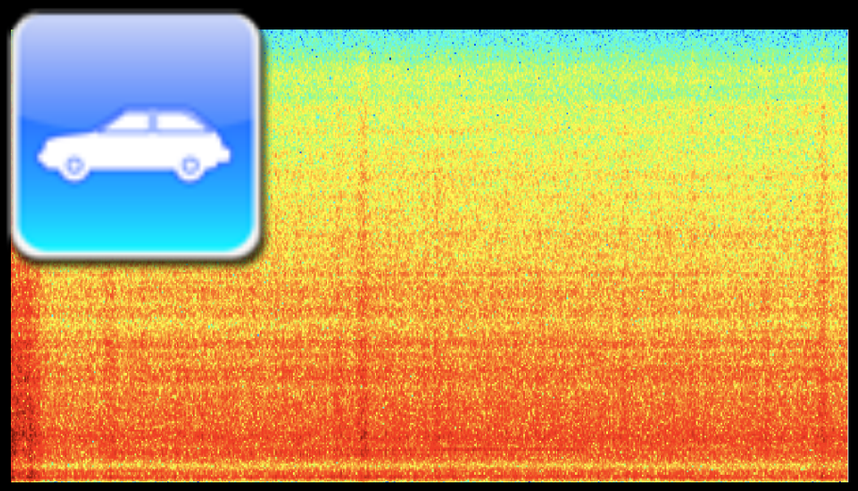
Coarse Category Classification

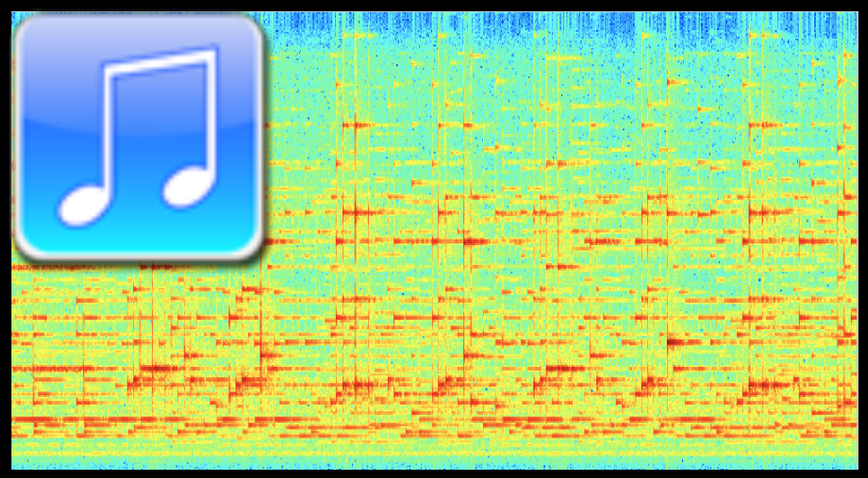


Intra-Category Classification

Discriminative Acoustic Features

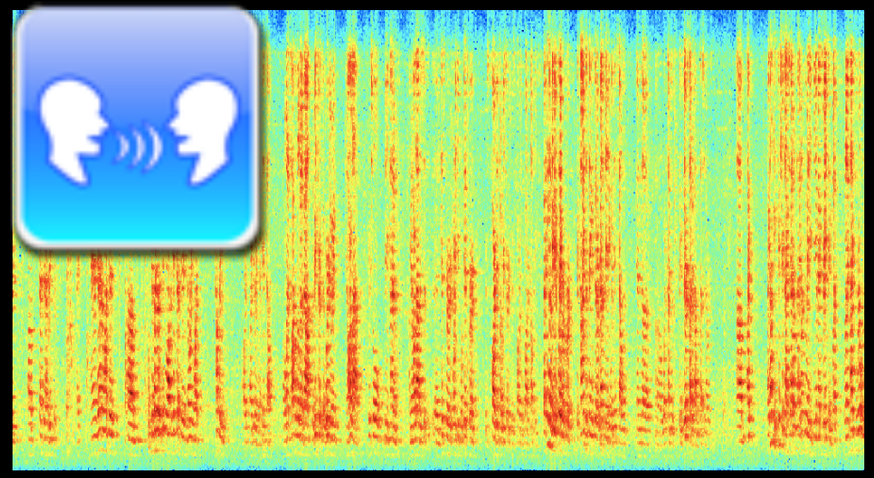


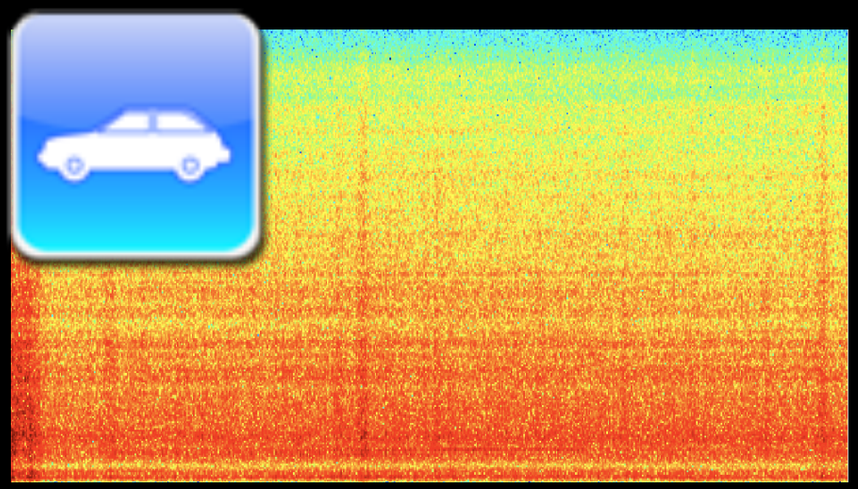


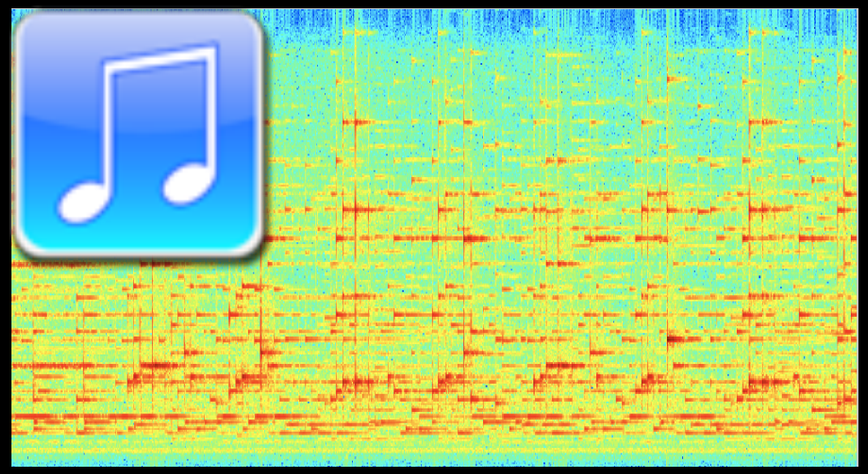


- Frequency Spectrum based features:
  - Robust to distance/muffling
  - Computationally very cheap after initial FFT

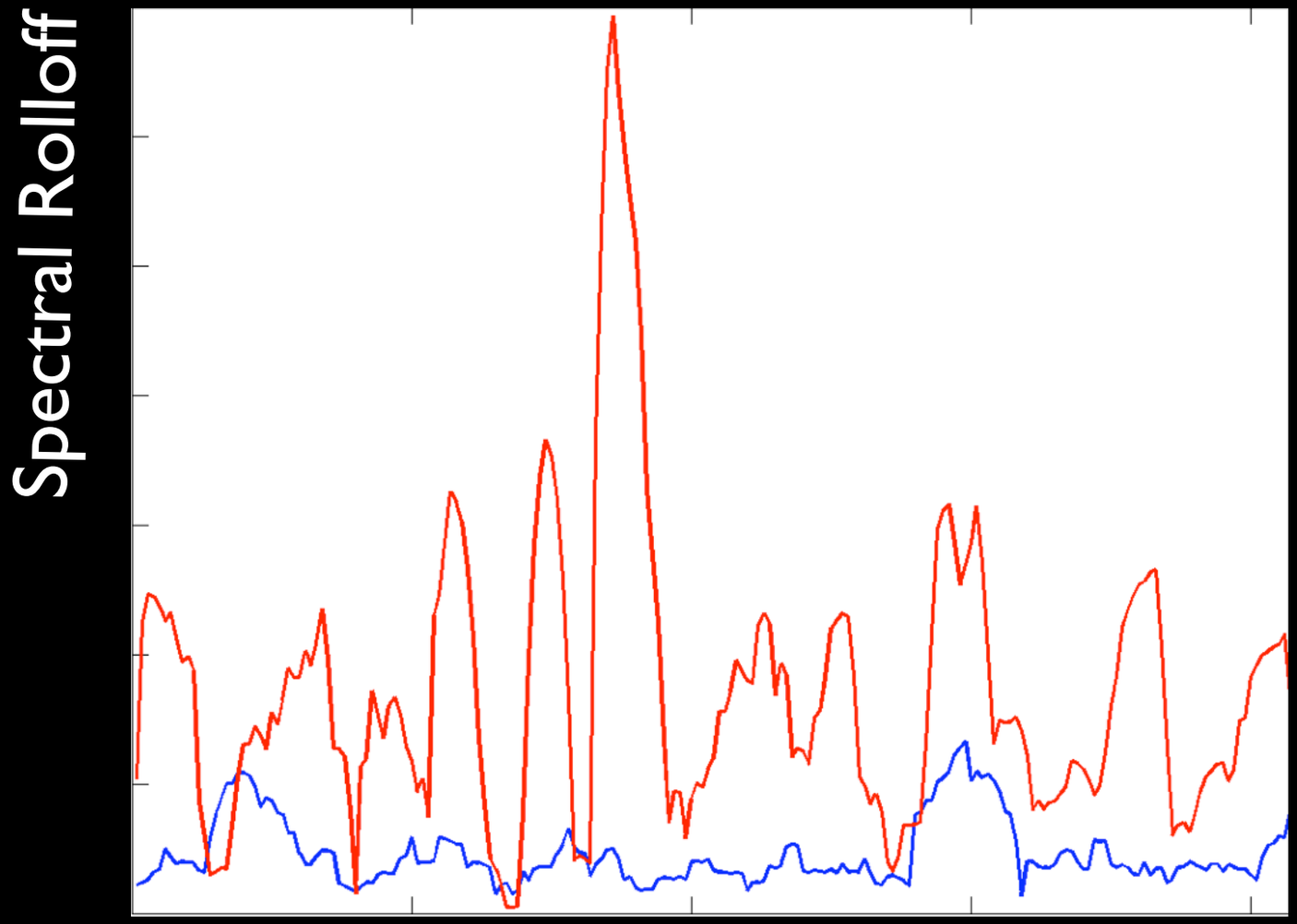
Discriminative Acoustic Features







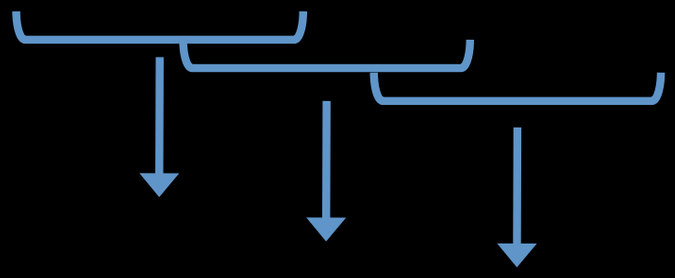
Spectral Rolloff for Music and Voice



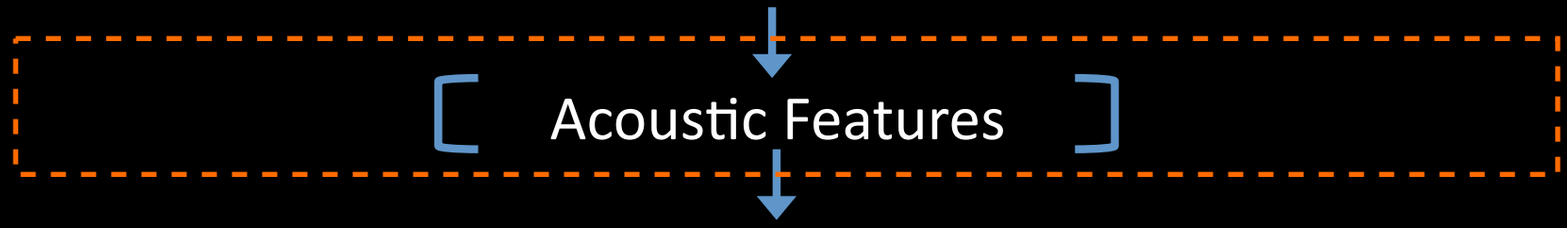
SoundSense Pipeline

Sound Waveform

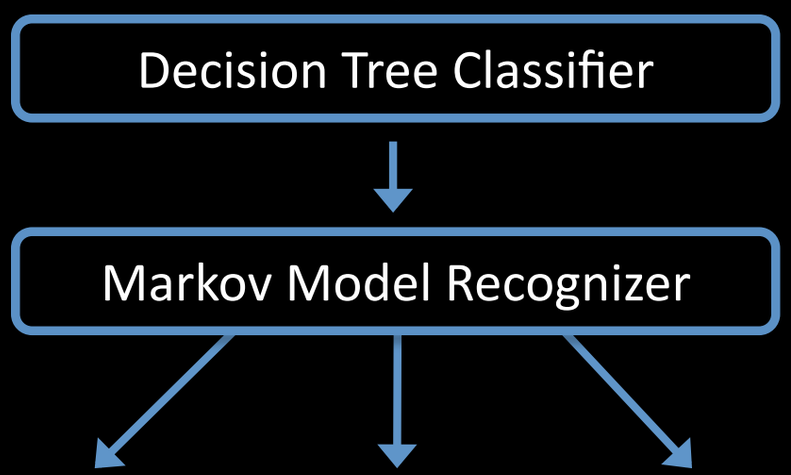
Framing



Feature Extraction



Coarse Category Classification

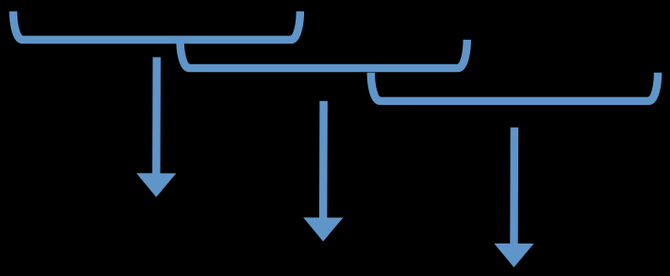


Intra-Category Classification

SoundSense Pipeline

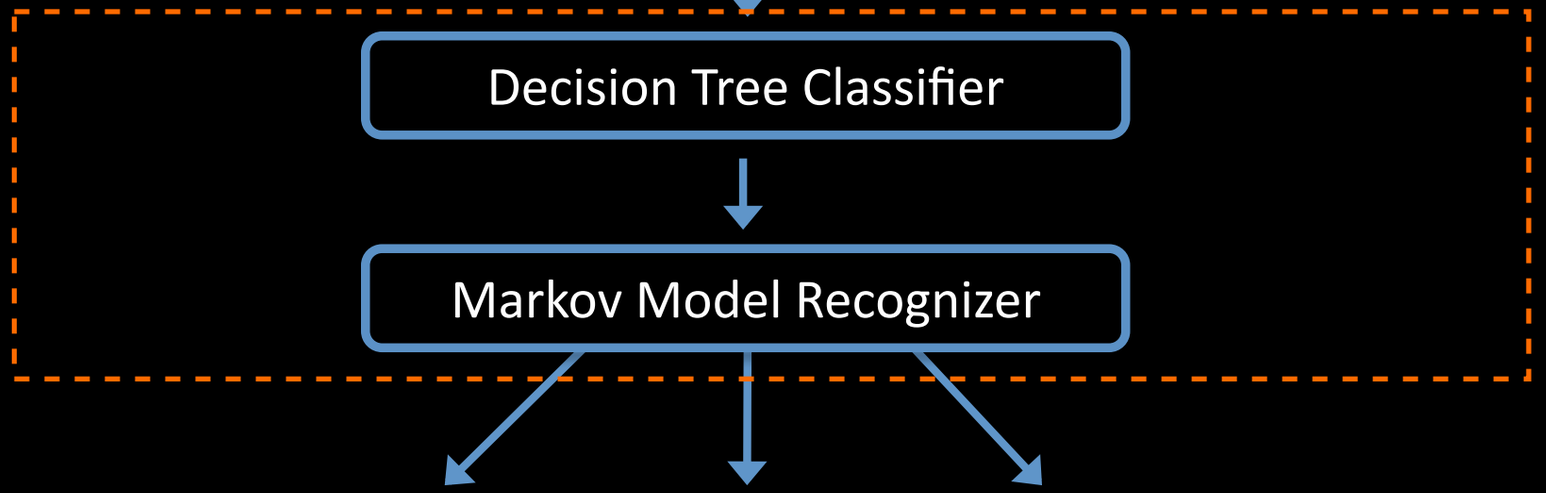
Sound Waveform

Framing



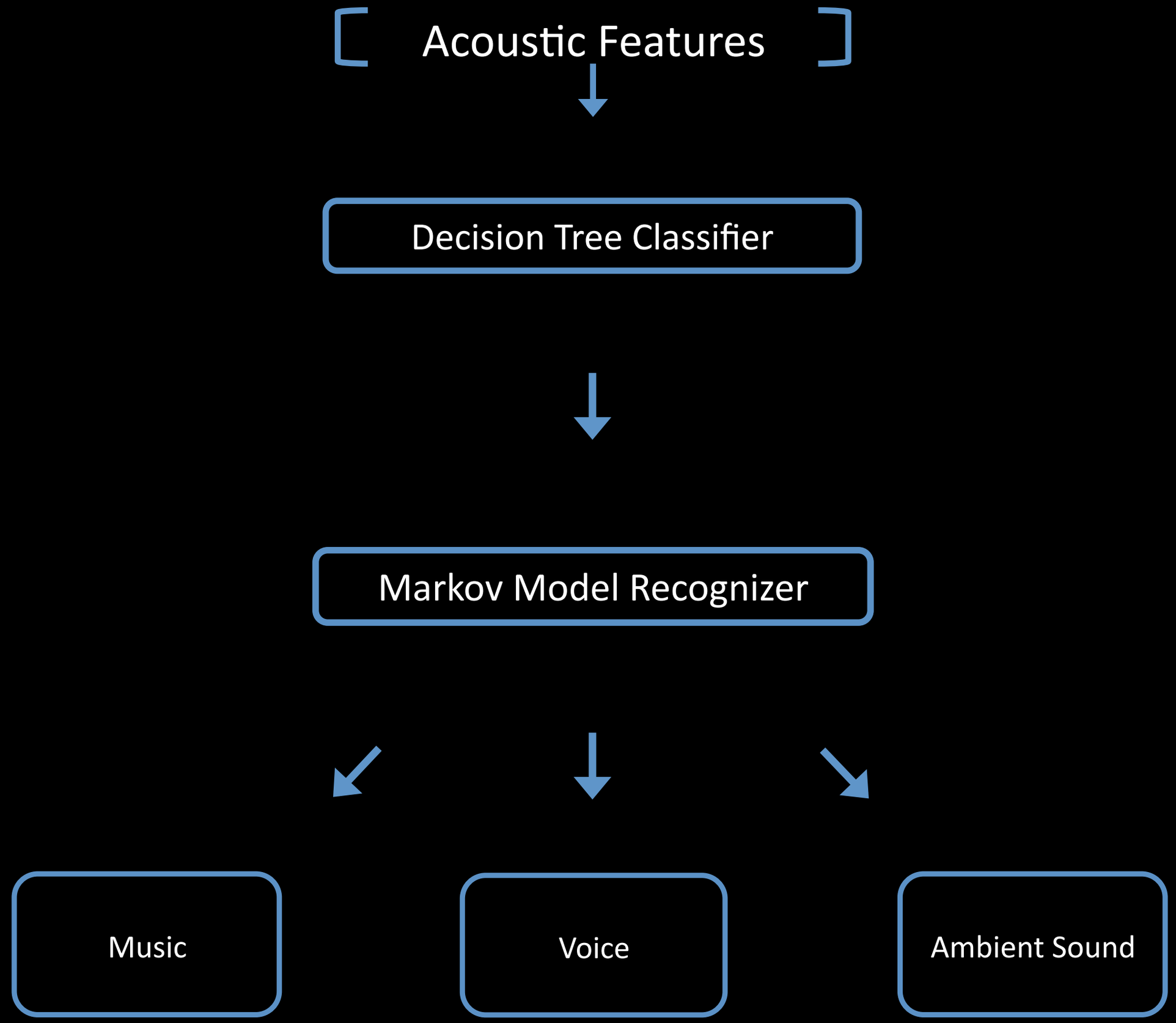
Feature Extraction

Coarse Category Classification



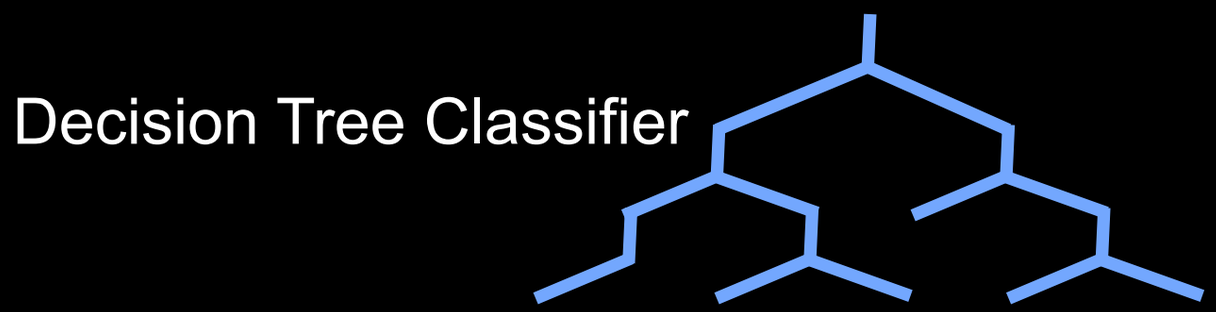
Intra-Category Classification

Coarse Category Classification



Coarse Category Classification

Acoustic Features



Markov Model Recognizer

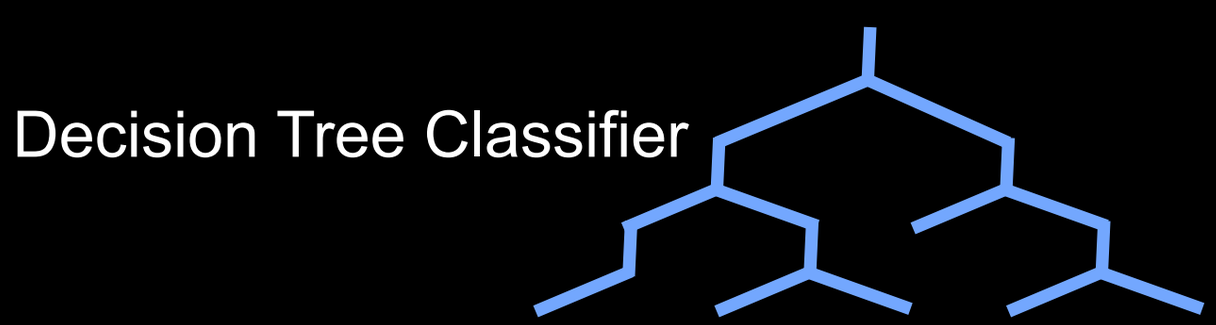
Music

Voice

Ambient Sound

Coarse Category Classification

Acoustic Features



Music, Music,Voice,Music,Voice, Music, Music,Voice

Markov Model Recognizer

Music

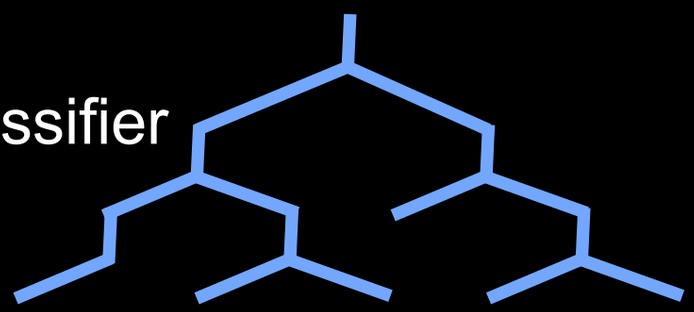
Voice

Ambient Sound

Coarse Category Classification

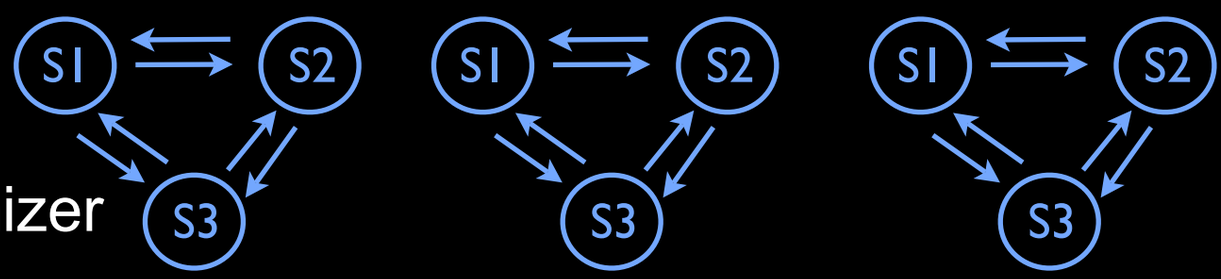
Acoustic Features

Decision Tree Classifier



Music, Music, Voice, Music, Voice, Music, Music, Voice

Markov Model Recognizer



Music

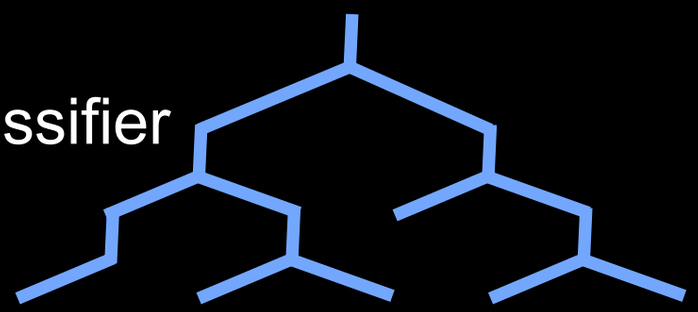
Voice

Ambient Sound

Coarse Category Classification

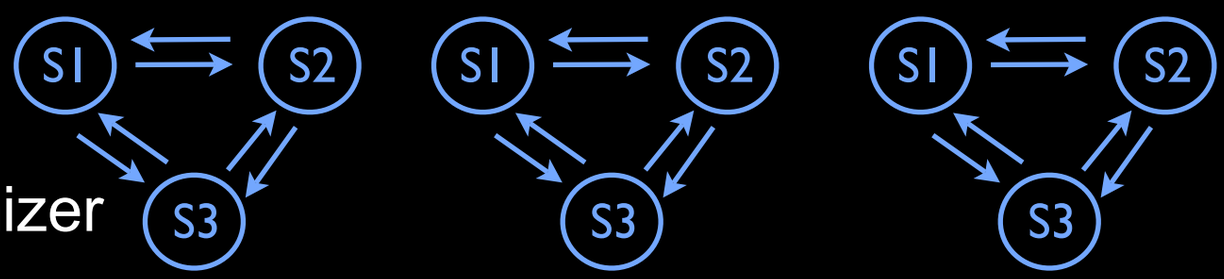
Acoustic Features

Decision Tree Classifier



Music, Music,Voice,Music,Voice, Music, Music,Voice

Markov Model Recognizer



Music, Music,Music,Music, Music, Music, Music,Music

Music

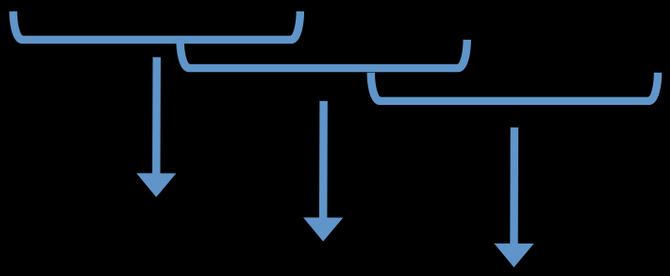
Voice

Ambient Sound

SoundSense Pipeline

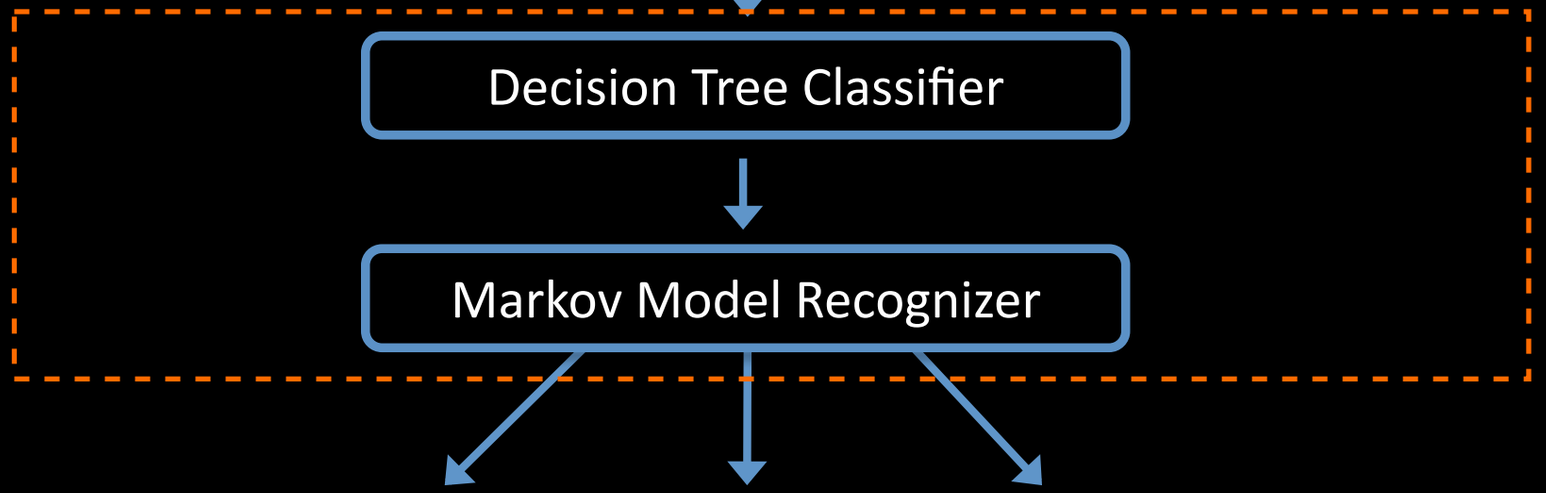
Sound Waveform

Framing



Feature Extraction

Coarse Category Classification

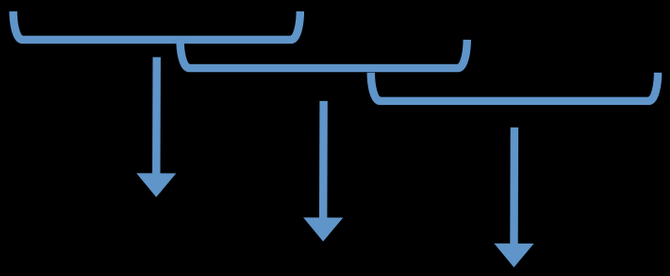


Intra-Category Classification

SoundSense Pipeline

Sound Waveform

Framing



Feature Extraction

Admission Control

Acoustic Features

Coarse Category Classification

Decision Tree Classifier

Markov Model Recognizer

Intra-Category Classification

Voice Analysis

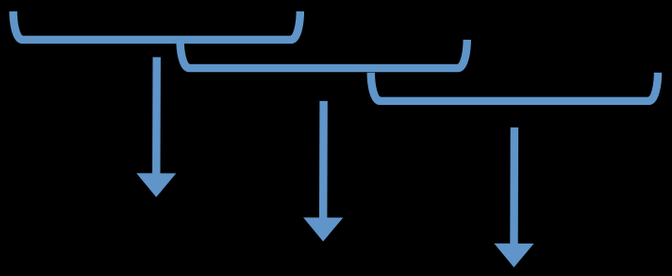
Music Analysis

Ambient Sound Learning

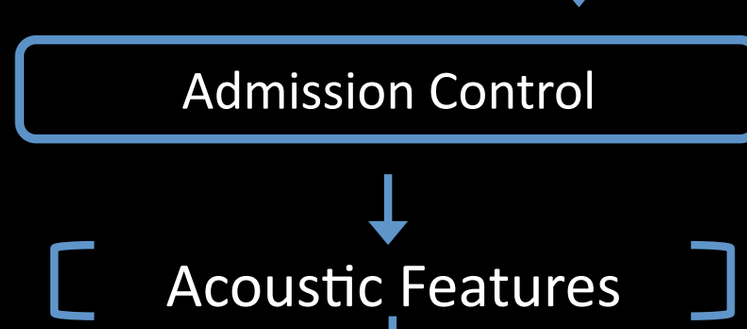
SoundSense Pipeline

Sound Waveform

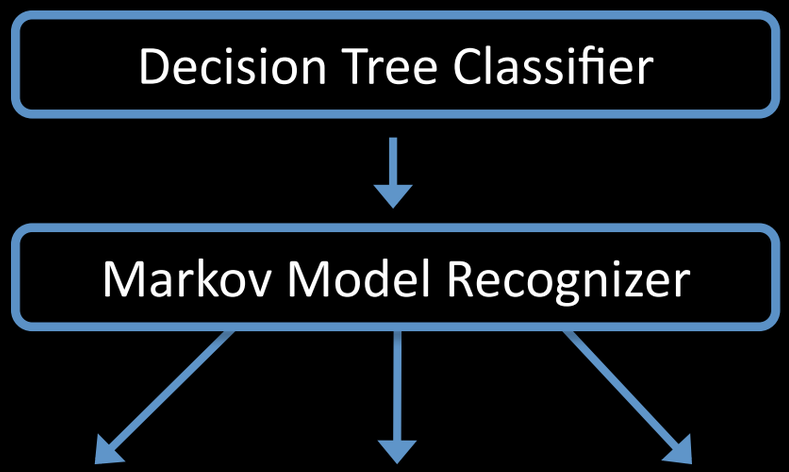
Framing



Feature Extraction



Coarse Category Classification



Intra-Category Classification

Significant sounds in the life of a user

- Unique to the individual
- Not practical to build supervised classifiers for all the potential significant sounds
- Examples:
  - Mystery Sound 1
  - Mystery Sound 2

Significant sounds in the life of a user

- Unique to the individual
- Not practical to build supervised classifiers for all the potential significant sounds
- Examples:
  - Mystery Sound 1
  - Mystery Sound 2

Significant sounds in the life of a user

- Unique to the individual
- Not practical to build supervised classifiers for all the potential significant sounds
- Examples:
  - Mystery Sound 1
  - Mystery Sound 2

Significant sounds in the life of a user

- Unique to the individual
- Not practical to build supervised classifiers for all the potential significant sounds
- Examples:
  - Car (turning signal)
  - Mystery Sound 2

Significant sounds in the life of a user

- Unique to the individual
- Not practical to build supervised classifiers for all the potential significant sounds
- Examples:
  - Car (turning signal)
  - Mystery Sound 2

Significant sounds in the life of a user

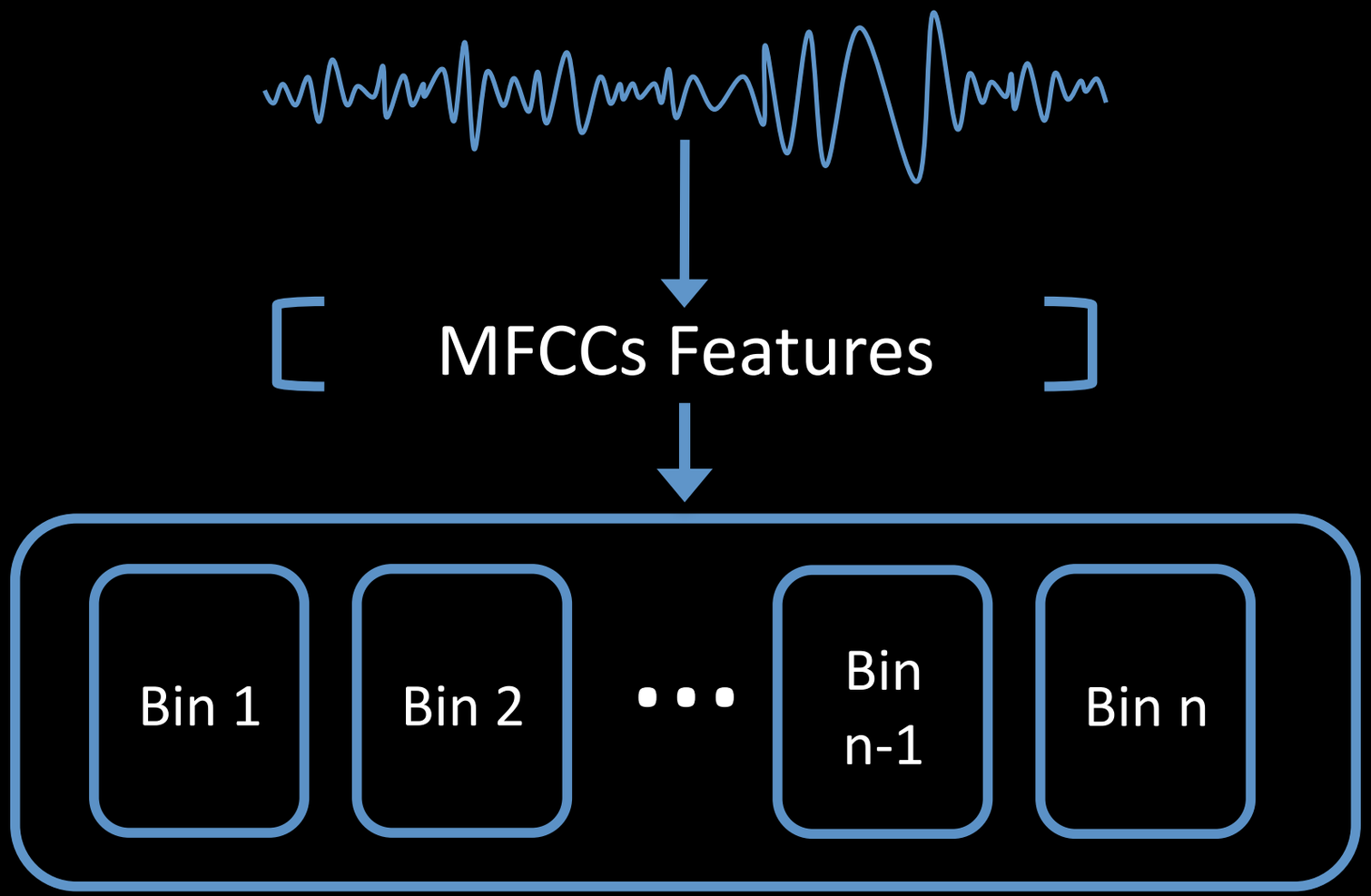
- Unique to the individual
- Not practical to build supervised classifiers for all the potential significant sounds
- Examples:
  - Car (turning signal)
  - Vacuum Cleaner

Our Approach

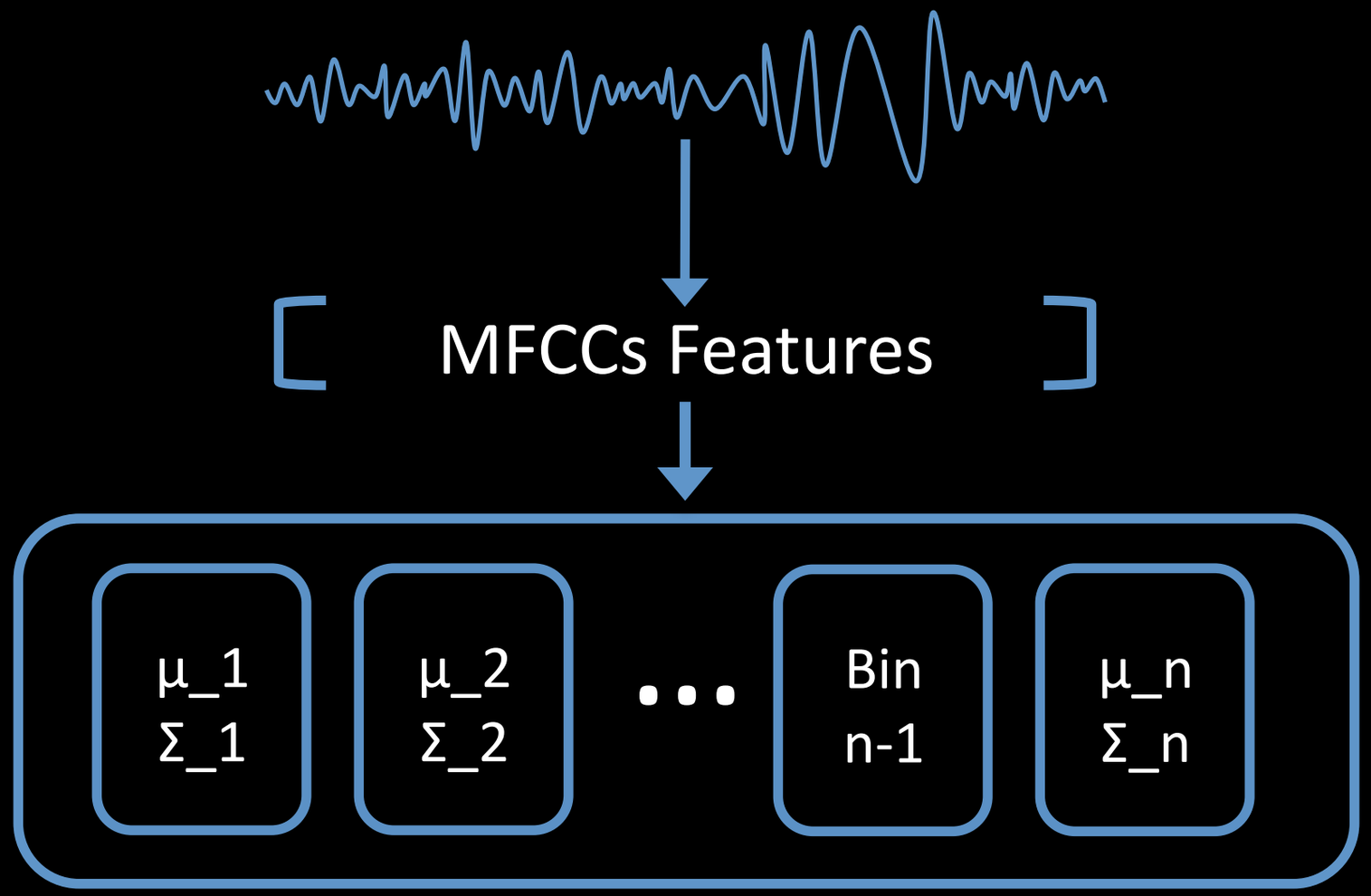
Incrementally build models of encountered sounds

Rank them by order of "importance"

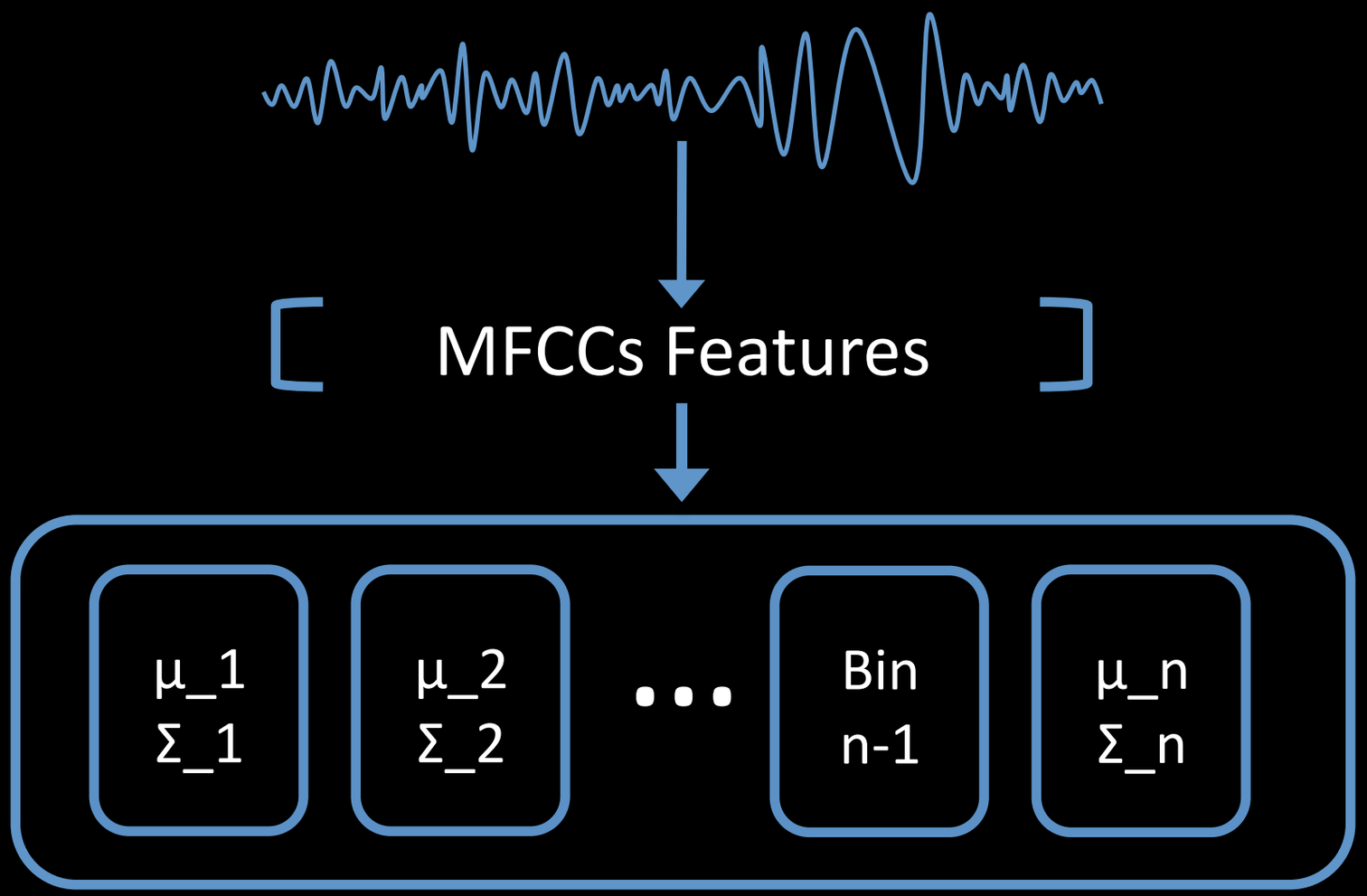
Human-in-the-loop labeling


MFCCs Features
Bin 1
Bin 2
Bin n-1
Bin n

Sounds modeled as multivariate Gaussians

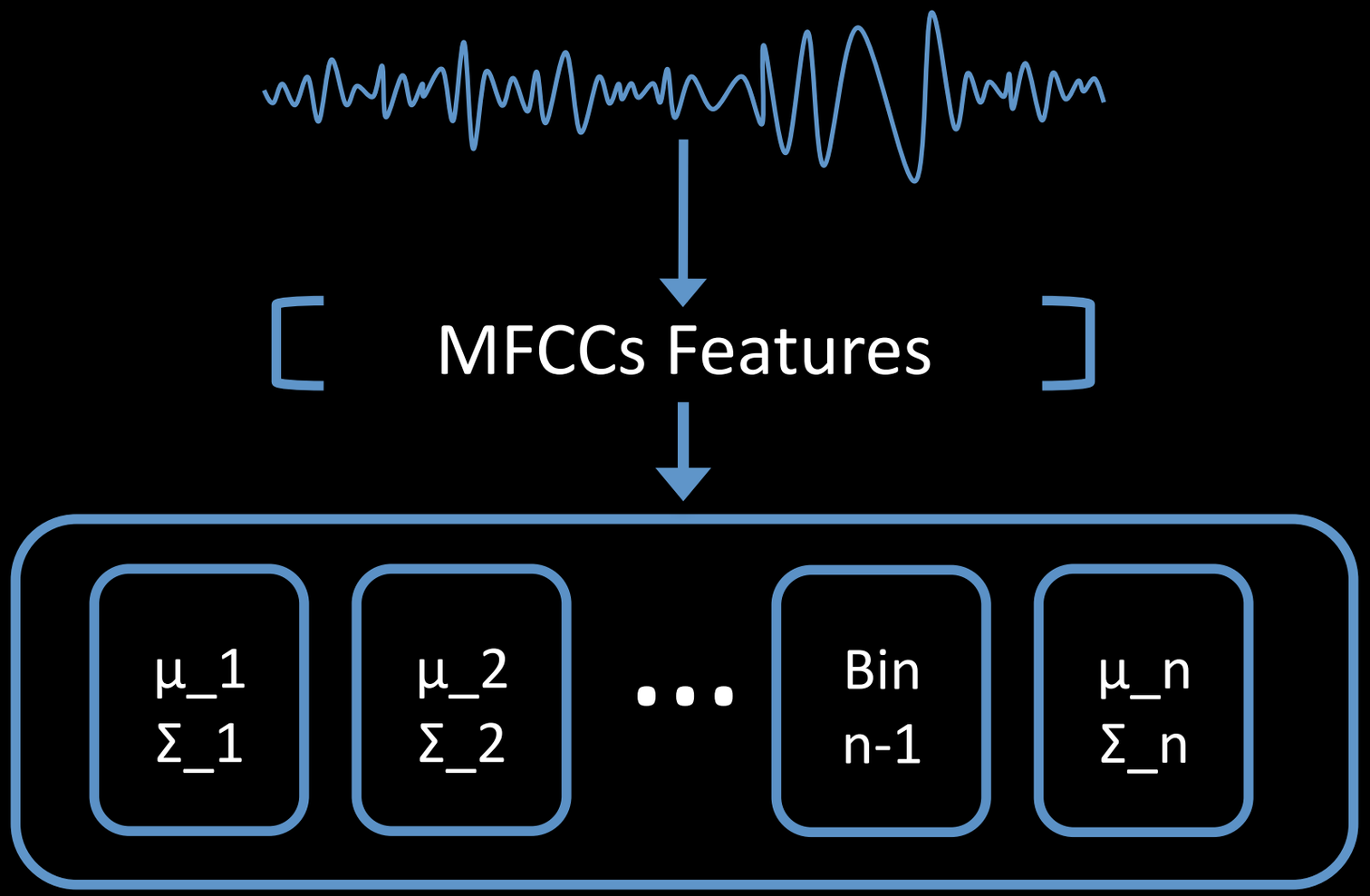


Sounds modeled as multivariate Gaussians



$$\text{Prob}(p_i = b) = \begin{cases} 1 & \text{if } i = \arg\max_b \mathcal{N}(\mathbf{m}_i; \mu_b, \Sigma_b) \\ & \text{and } \mathcal{N}(m_i; \mu_b, \Sigma_b) > \epsilon_{\text{THRESHOLD}} \\ 0 & \text{otherwise} \end{cases}$$

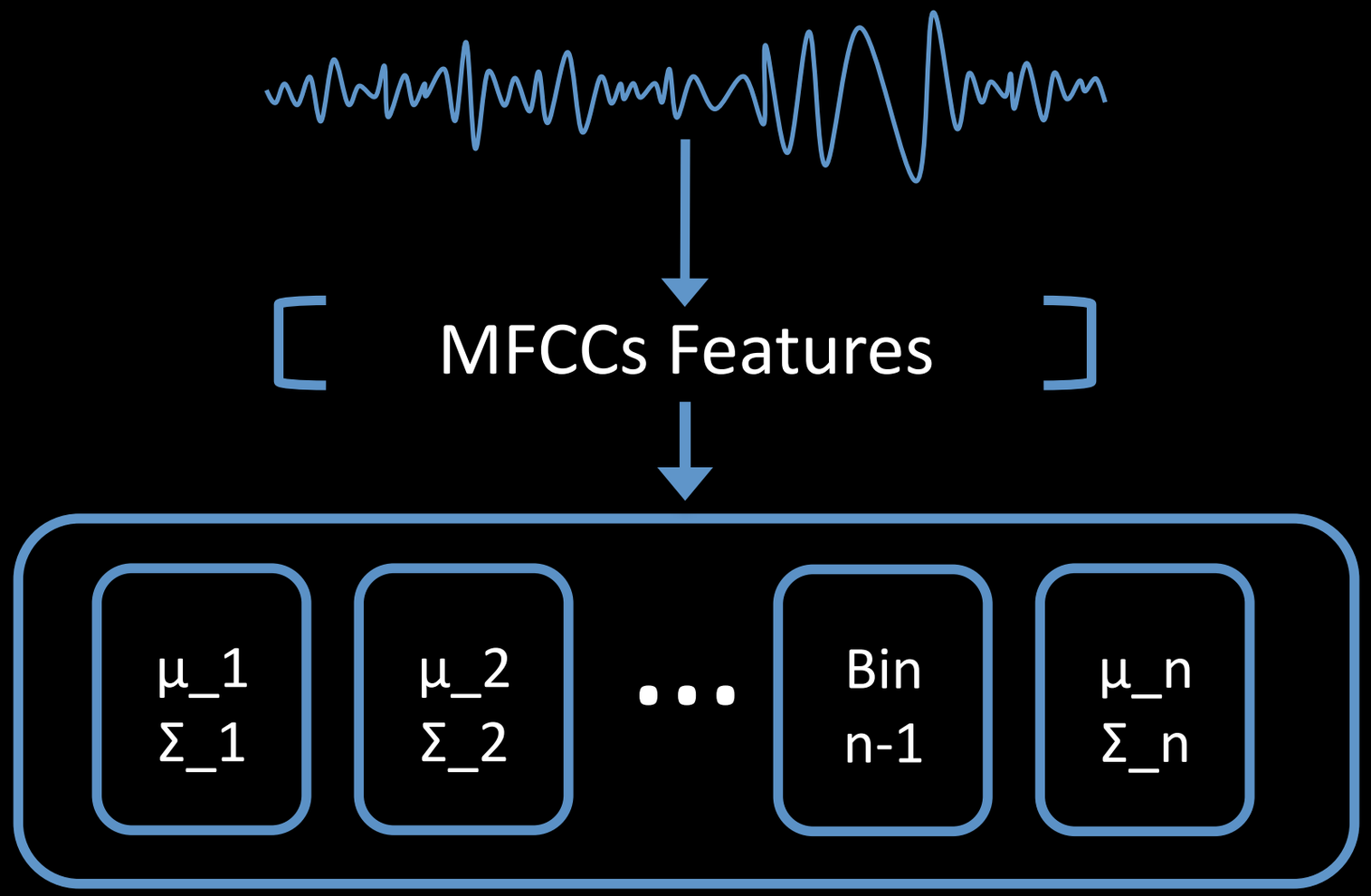
Sounds modeled as multivariate Gaussians



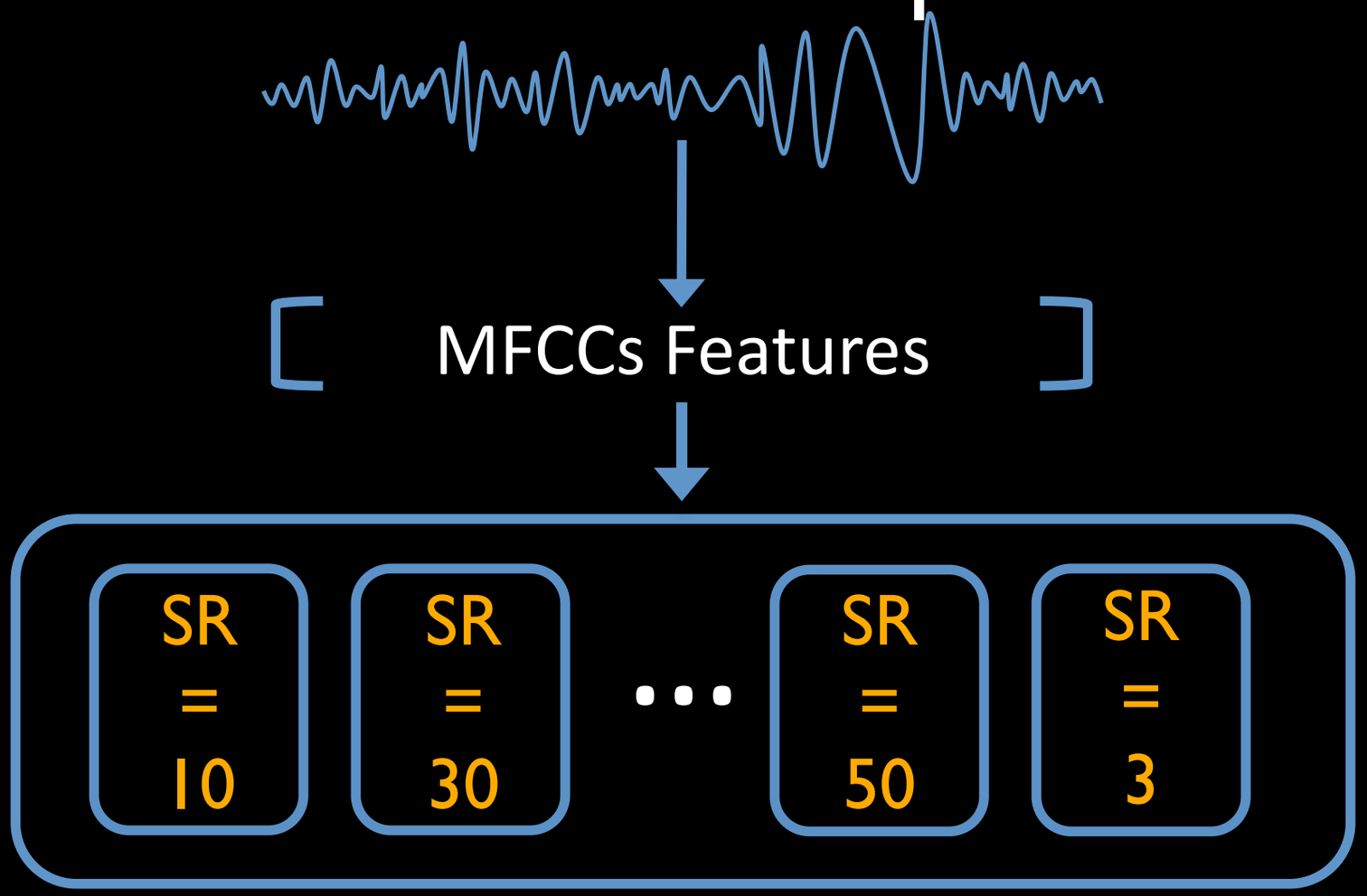
$$\mu_b' = \frac{\sum_{i=1}^{t} \text{Prob}(p_i = b)\mathbf{w}_i}{\sum_{i=1}^{t} \text{Prob}(p_i = b)} = \frac{1}{t}(\mu_b(t-1) + \mathbf{w}_{j,\text{Prob}(p_j=b)=1})$$

$$\Sigma_b' = \frac{\sum_{i=t-P}^{t} \text{Prob}(p_i = b)(\mathbf{w}_i - \mu_b')(\mathbf{w}_i - \mu_b')^T}{\sum_{i=1}^{m} \text{Prob}(p_i = b)}$$

Sounds modeled as multivariate Gaussians



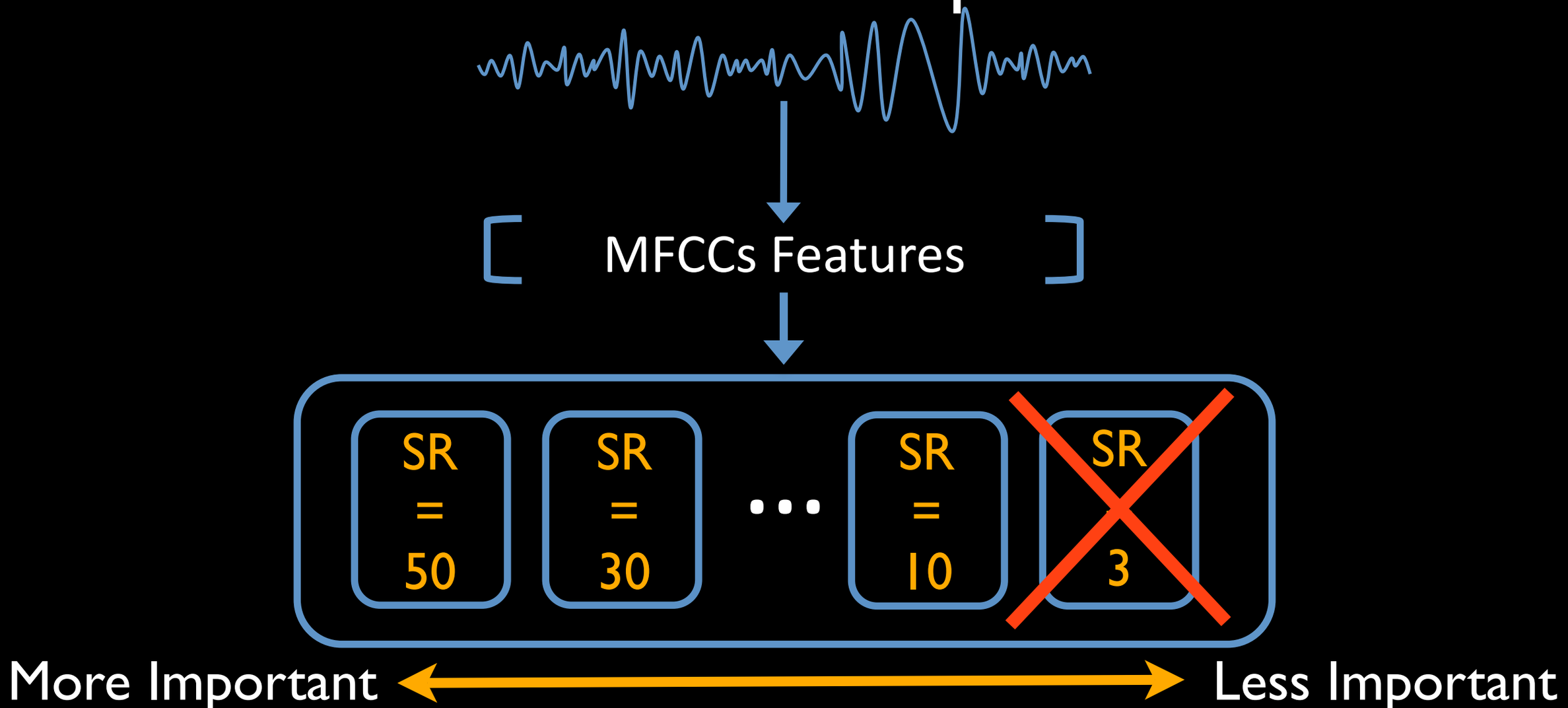
Sound Rank estimates the “importance” a sound



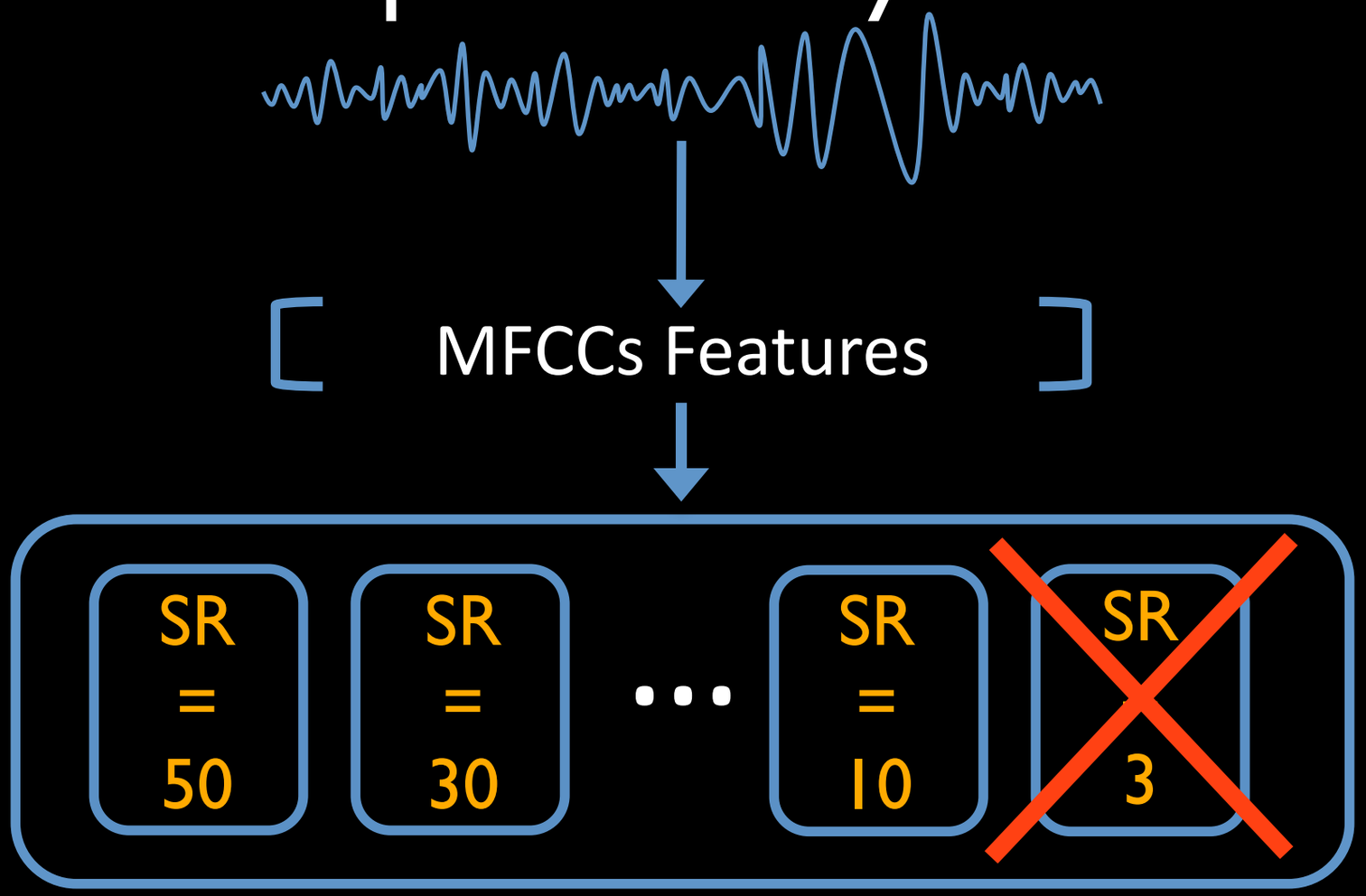
Sound Rank based on:

- Duration of the event
- Frequency of the encounter

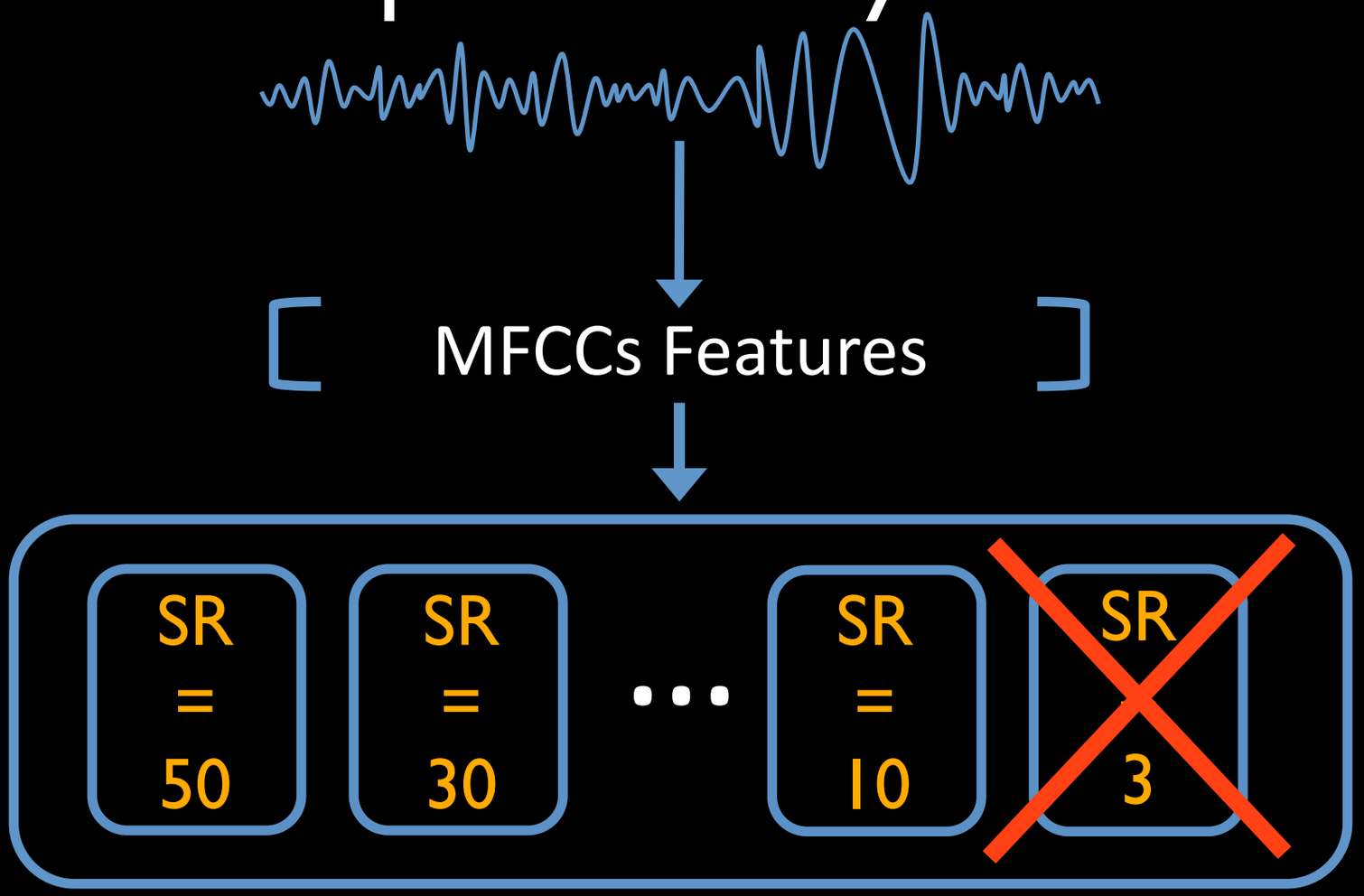
Sound Rank estimates the "importance" a sound

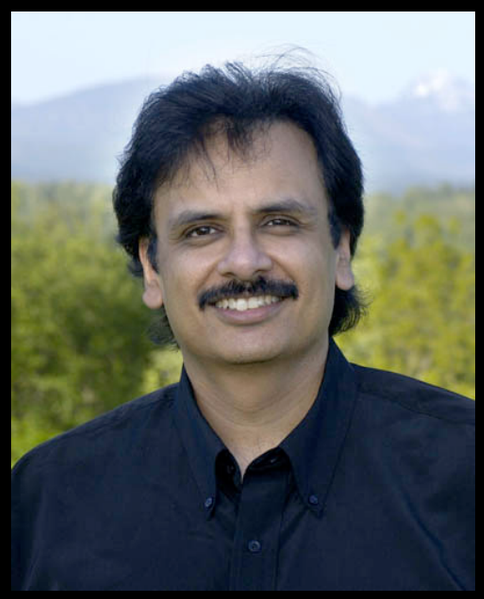


Human-in-the-loop to identify and validate sounds

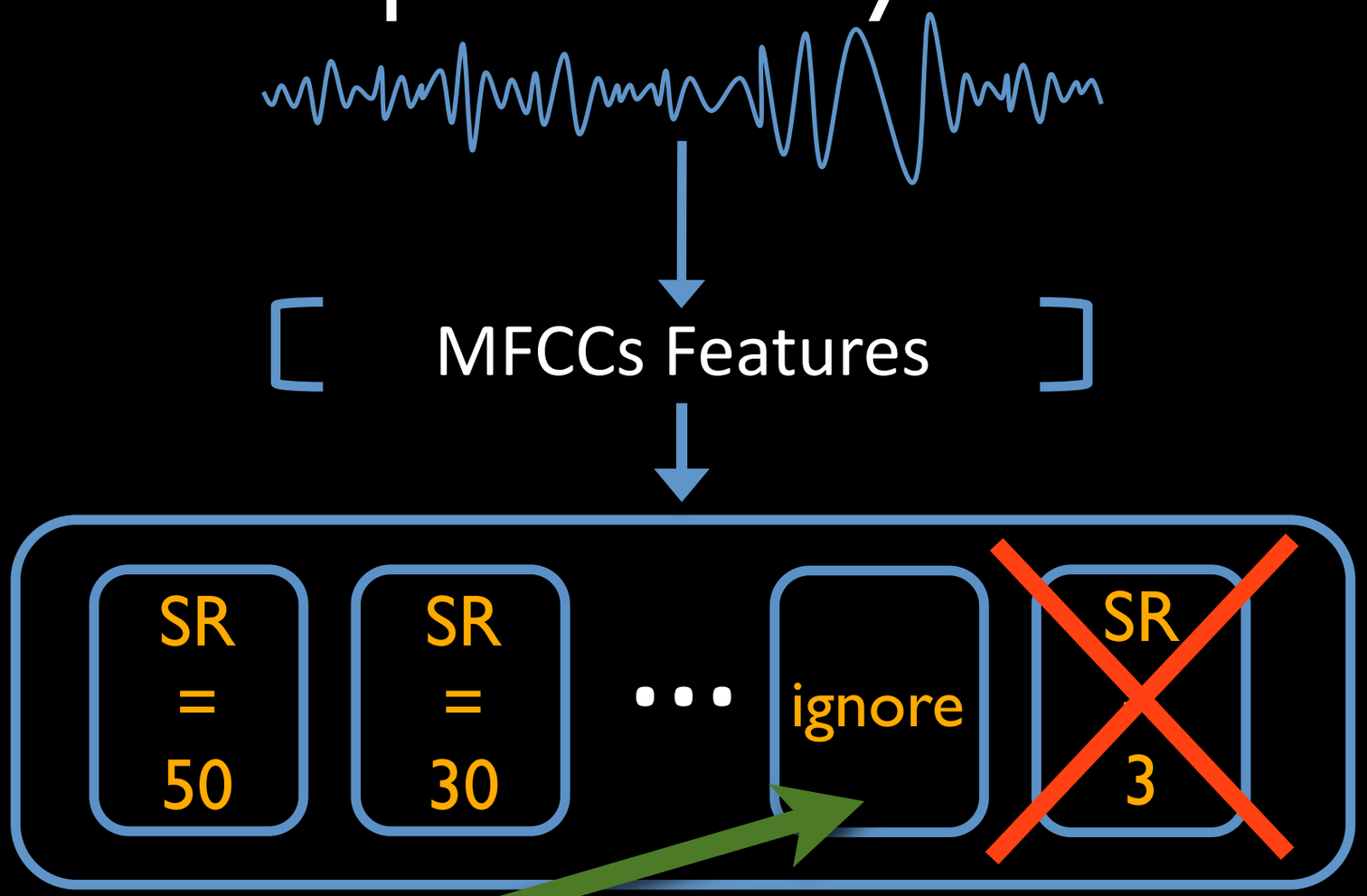


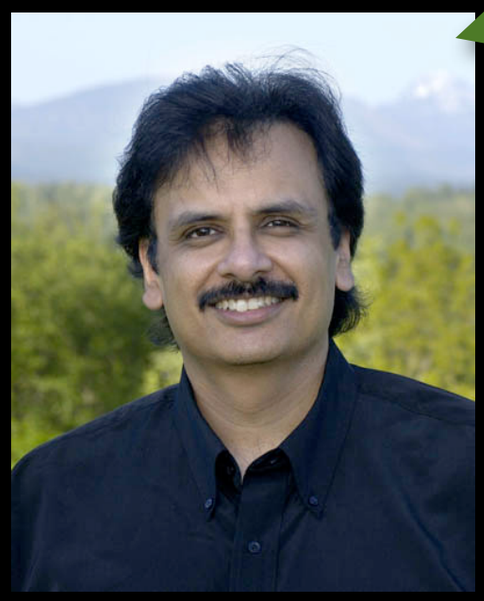
Human-in-the-loop to identify and validate sounds



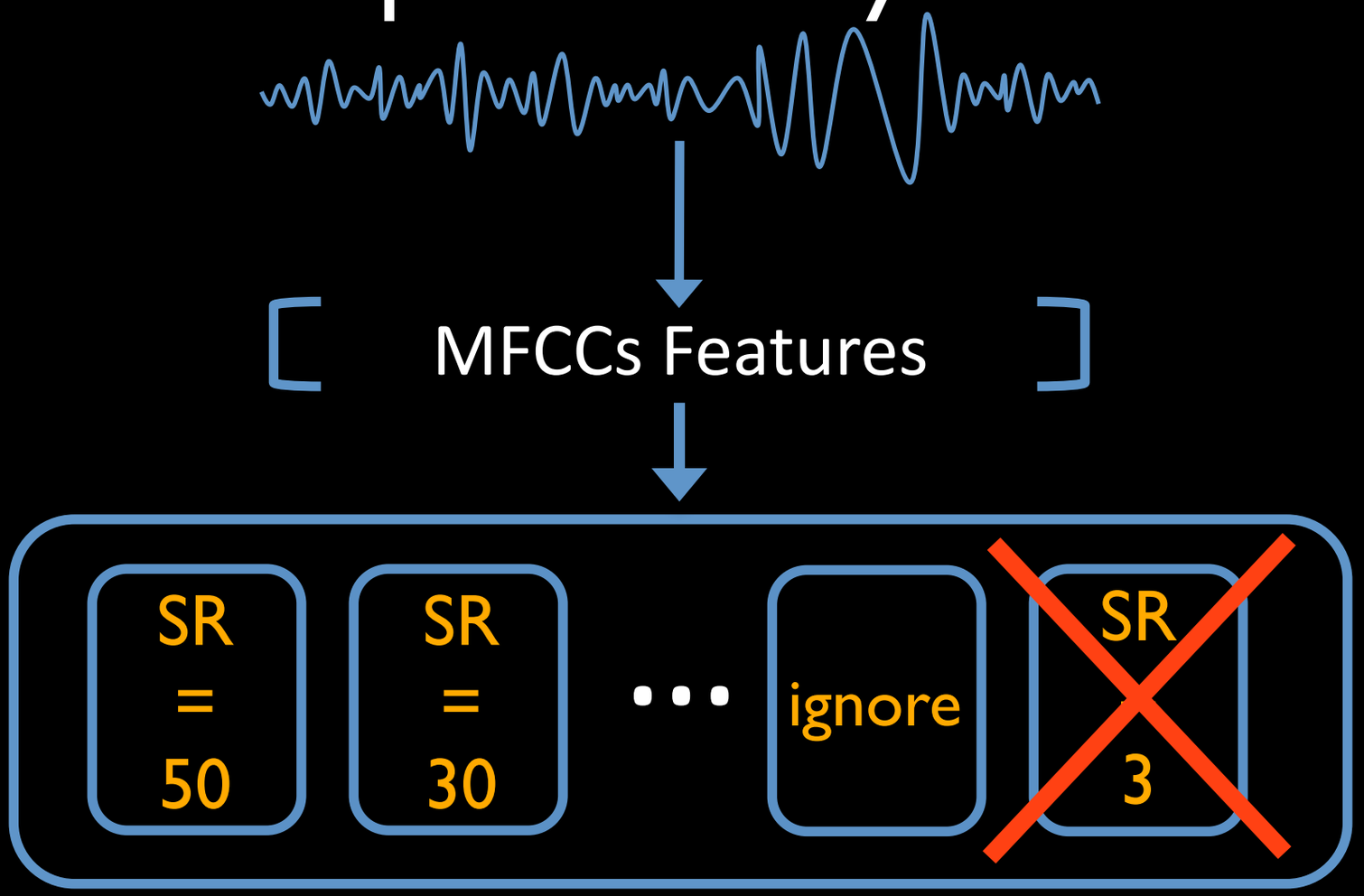


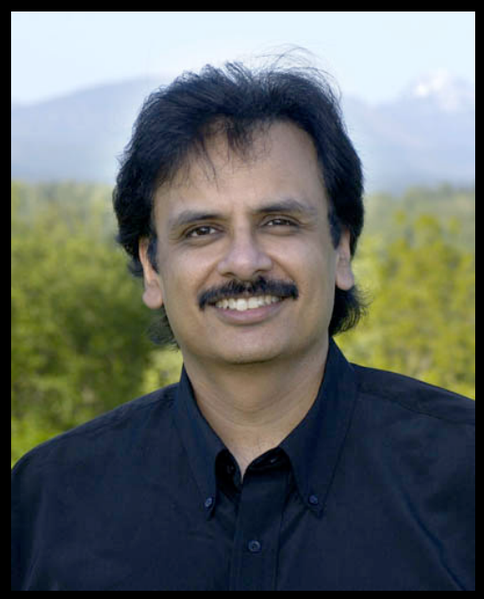
Human-in-the-loop to identify and validate sounds



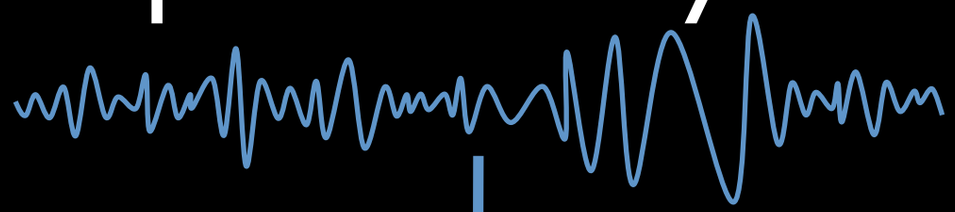


Human-in-the-loop to identify and validate sounds





Human-in-the-loop to identify and validate sounds



MFCCs Features

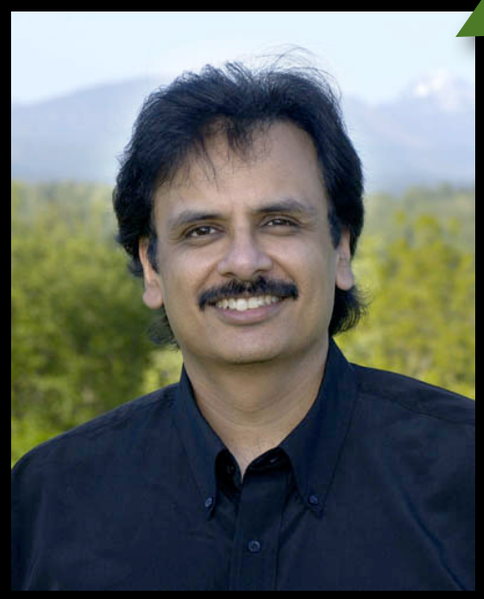
SR = 50

hide

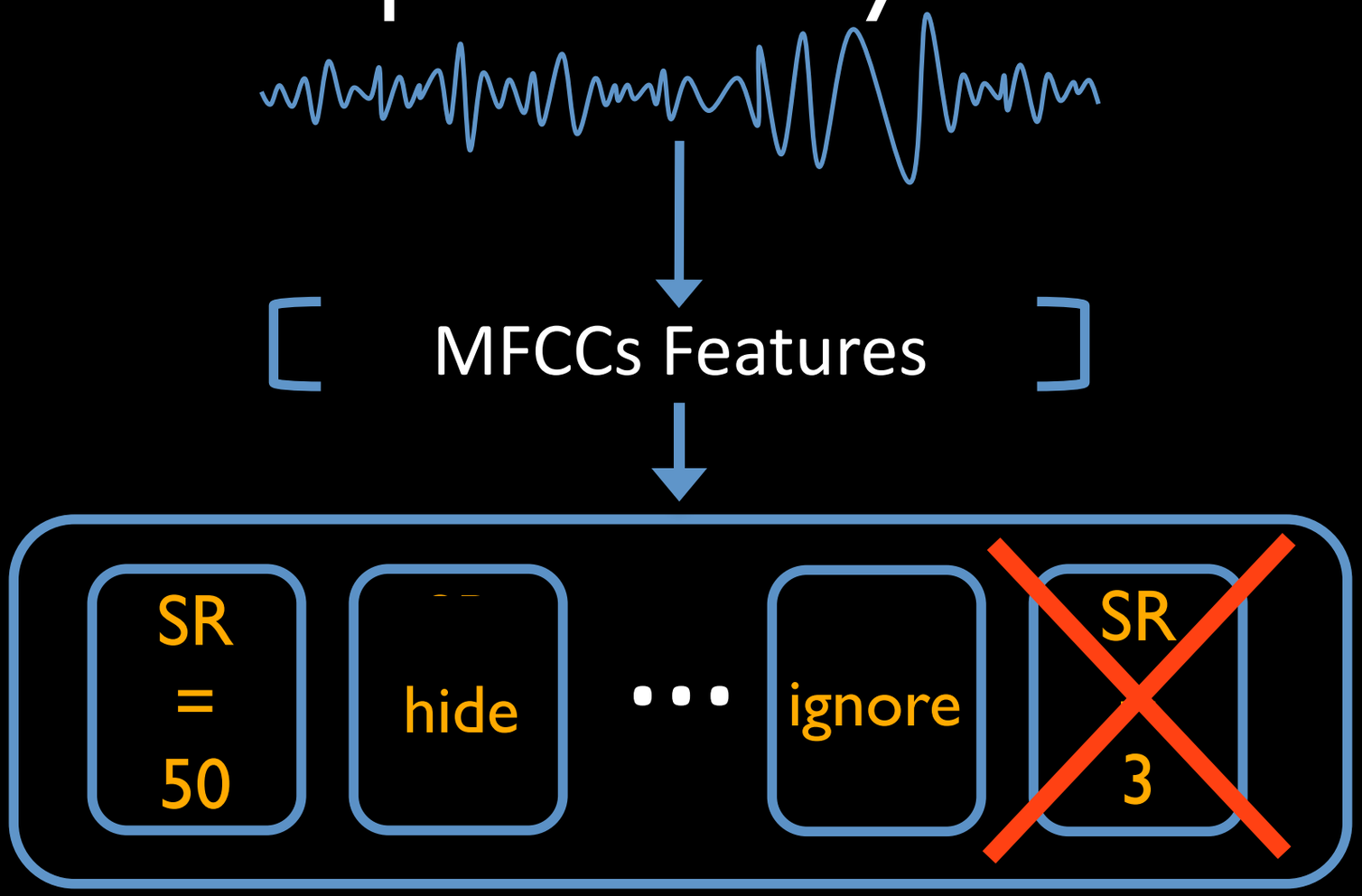
• • •

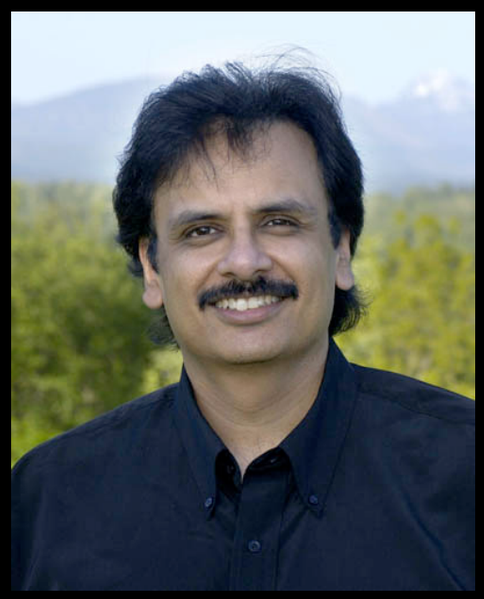
ignore

SR = 3



Human-in-the-loop to identify and validate sounds





Human-in-the-loop to identify and validate sounds

MFCCs Features

label

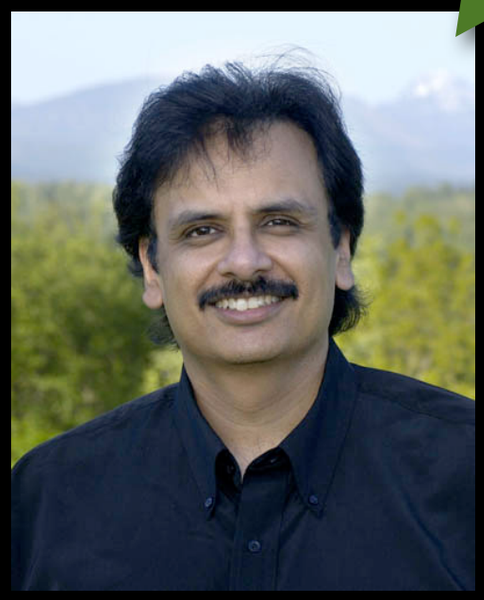
hide

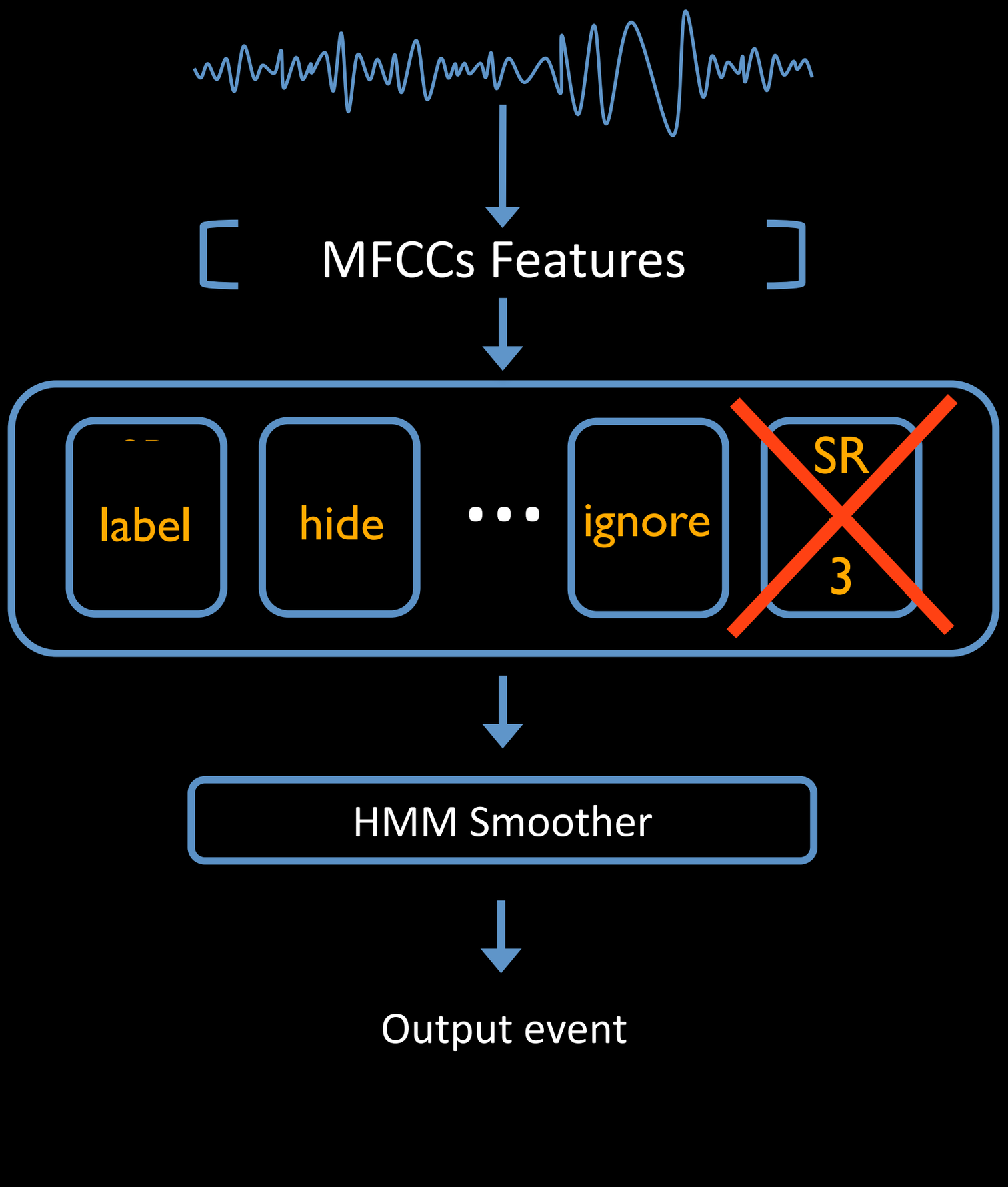
• • •

ignore

SR

3




MFCCs Features
label
hide
ignore
SR
3
HMM Smoother
Output event

evaluation

iPhone implementation

- Jail broken iPhone
- C, C++, Objective C
- 300 KB application footprint

SoundSense implementation has low levels of resource consumption on the iPhone

	cpu usage range	memory usage range
GUI only	< 1%	2.79 MB
silence	**1 - 5 %**	**3.75 MB**
music/speech	8% - 20%	3.80 MB
ambient sound	11% - 22%	3.80 MB ~ 5 MB

SoundSense implementation has low levels of resource consumption on the iPhone

	cpu usage range	memory usage range
GUI only	< 1%	2.79 MB
silence	1 - 5 %	3.75 MB
music/speech	8% - 20%	3.80 MB
ambient sound	**11% - 22%**	**3.80 MB ~ 5 MB**

Coarse Classification using Decision Tree only

	ambient sound	music	voice
ambient sound	0.91	0.06	0.02
music	0.13	0.81	0.05
voice	0.06	0.14	0.78

Coarse Classification using Decision Tree only

	ambient sound	music	voice
ambient sound	**0.91**	0.06	0.02
music	0.13	0.81	0.05
voice	0.06	0.14	0.78

Coarse Classification using Decision Tree only

	ambient sound	music	voice
ambient sound	0.91	0.06	0.02
music	0.13	**0.81**	0.05
voice	0.06	0.14	**0.78**

Coarse Classification using Decision Tree and Markov Model smoothing

	ambient sound	music	voice
ambient sound	0.94	0.04	0.01
music	0.04	**0.92**	0.04
voice	0.03	0.07	**0.90**

Unsupervised Adaptive Classifier Learning New Sound Classes

Sound Rank	**Event**
1st	Phone rubbing against clothes
2nd	Highway driving
3rd	Walking outdoors
4th	Driving w/ acceleration
5th	In-town driving
6th	Walking indoors
7th	Taking out trash

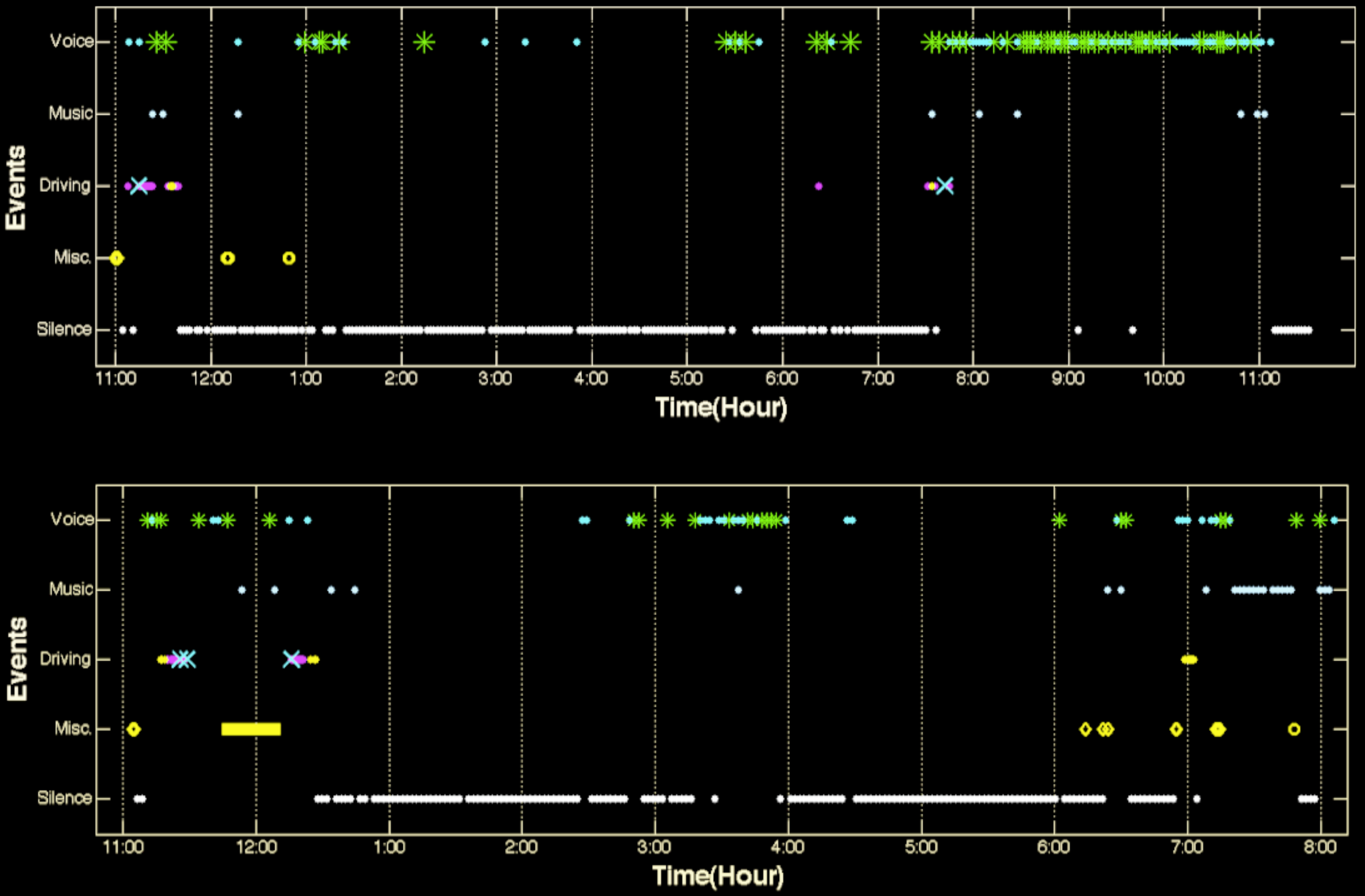
Unsupervised Adaptive Classifier
Learning New Sound Classes

Sound Rank	**Event**
1st	Phone rubbing against clothes
2nd	**Highway driving**
3rd	Walking outdoors
4th	**Driving w/ acceleration**
5th	**In-town driving**
6th	Walking indoors
7th	Taking out trash

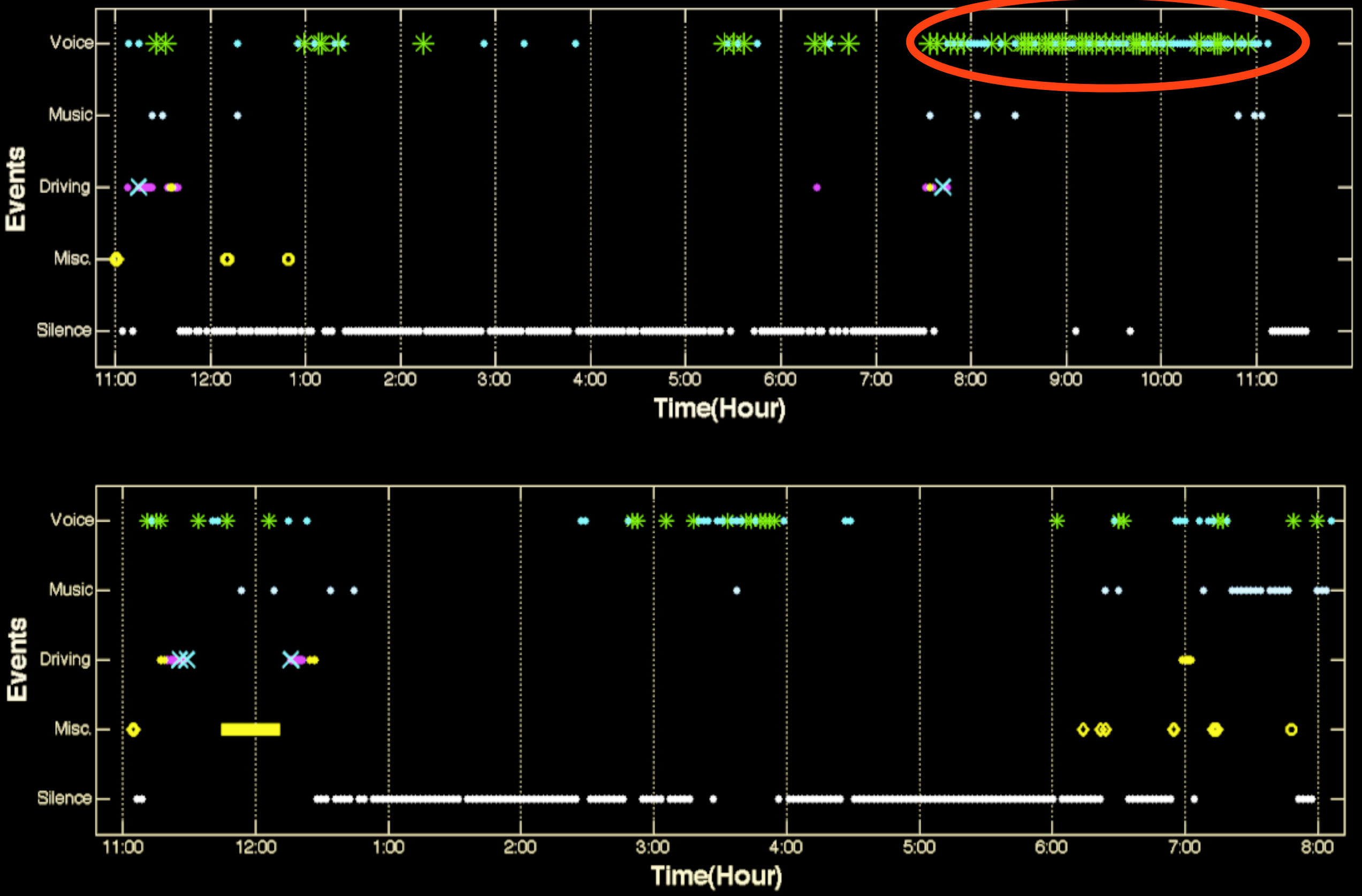
Unsupervised Adaptive Classifier Learning New Sound Classes

Sound Rank	**Event**
1st	**Phone rubbing against clothes**
2nd	Highway driving
3rd	Walking outdoors
4th	Driving w/ acceleration
5th	In-town driving
6th	Walking indoors
7th	Taking out trash

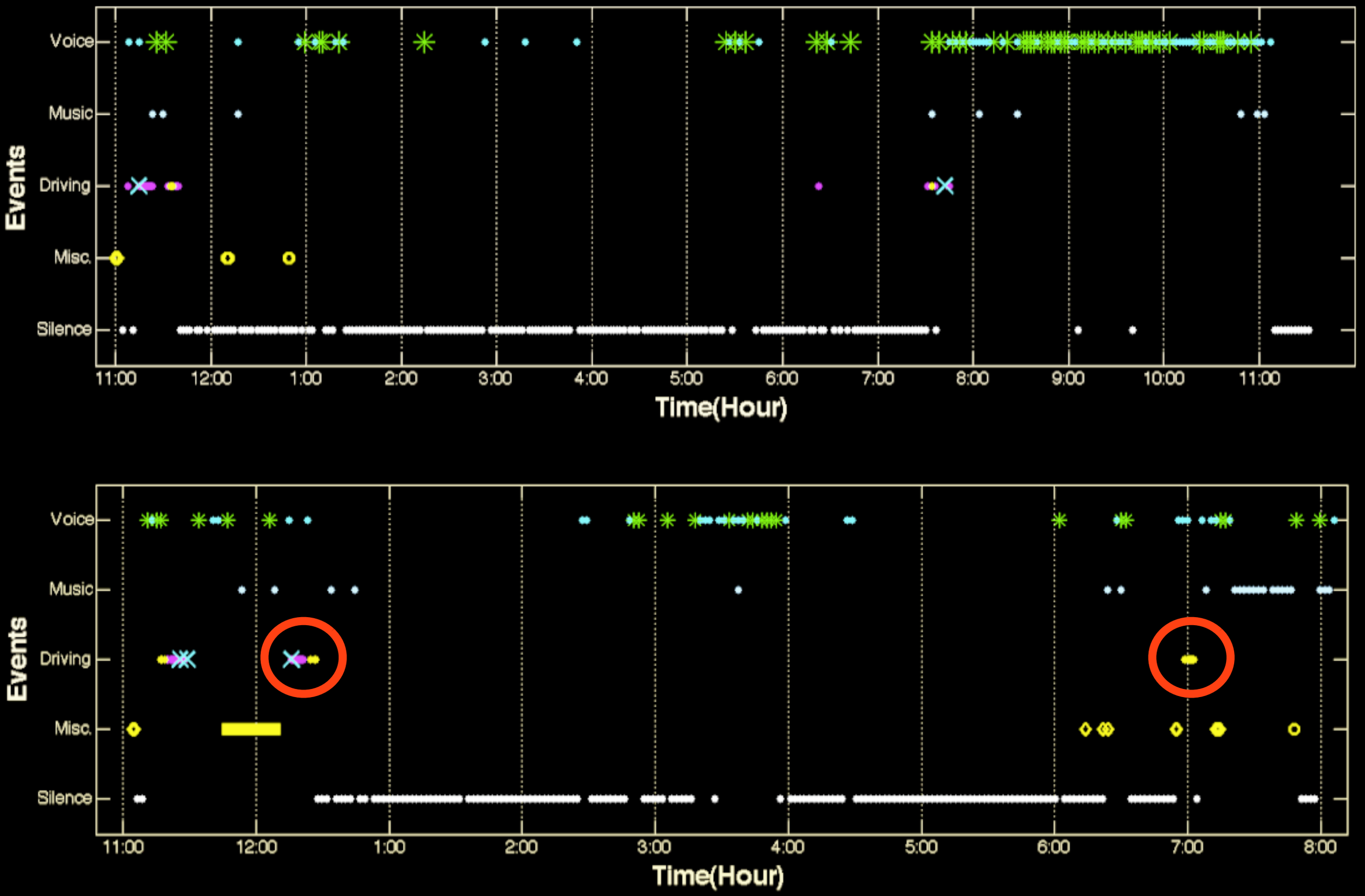
Sound Sense Diary: Discovering Life Patterns



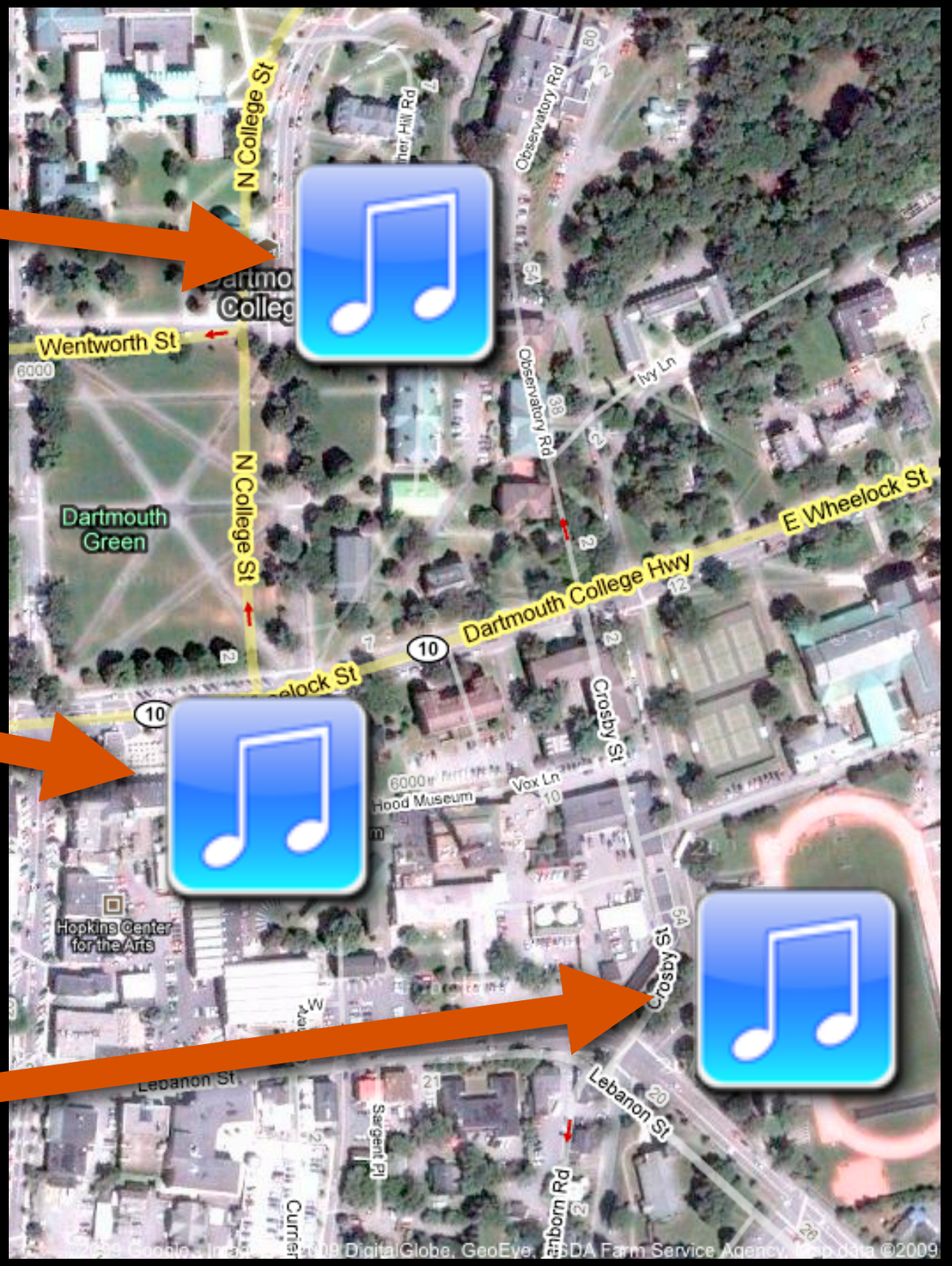
Sound Sense Diary: Discovering Life Patterns



Sound Sense Diary: Discovering Life Patterns



Music Tagger: Building Apps with SoundSense



Related Work

- recognition/audio scene analysis using resource constrained devices
  - e.g., Clarkson et al. ’98, Smith et al. ‘06
- specific types of sound analysis:
  - speech recognition, music genre classification, speaker identification
- sensing with mobile phones:
  - UCLA, UIUC, Nokia, Microsoft, Motorola, Intel, UW, Duke

Summary

- SoundSense addresses the needs of sound classification using mobile phones:
  - Privacy, Energy Efficiency & robustness to Phone Context
  - Scalability via unsupervised learning of everyday sounds
- Future Work - Improving Ambient Sound Learning
  - Revising Sound Rank
  - Coping with fragmented learnt classes

Thanks for listening

Questions?