USC University of Southern California

USC Viterbi School of Engineering

Ming Hsieh Department of Electrical Engineering

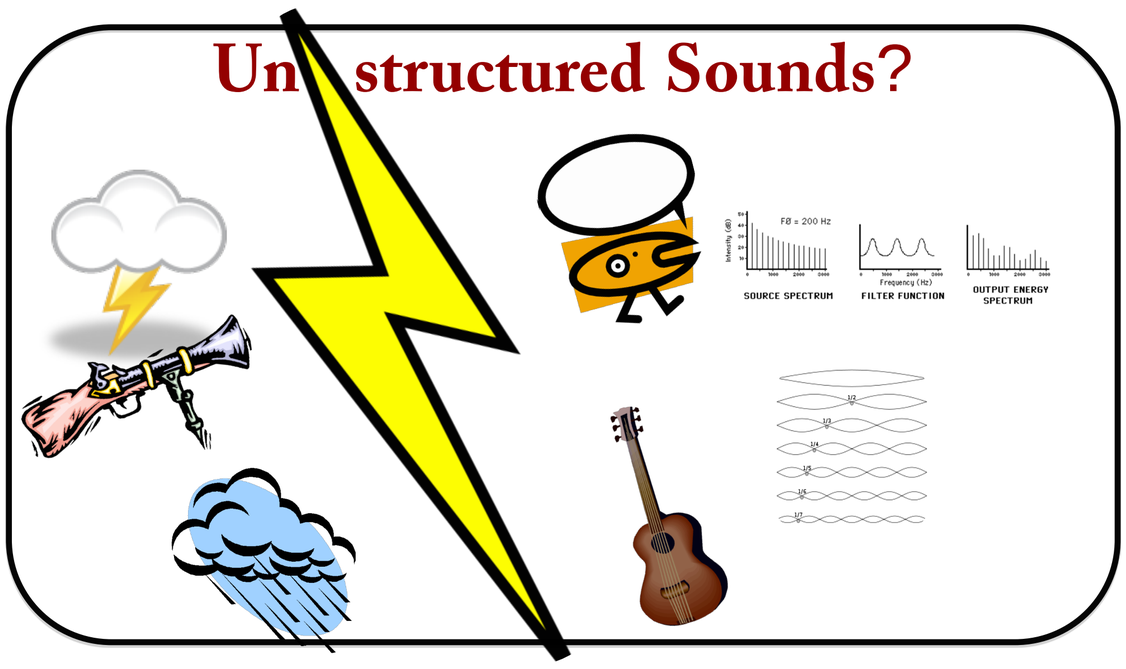
On Unstructured Audio Modeling with Statistical Analysis and Test

Sachin Chachada and C.-C. Jay Kuo

Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering

University of Southern California

Unstructured Sounds?



Simple Family of Classifiers (SFC)

$$p(c_k|\mathbf{x}) = p^{1-N}(c_k) \prod_{0 \le n \le N-1} p(c_k|\mathbf{x}_n)$$

Naïve Bayes Model

SVM

GMM LDA

Simple to train, low complexity, mathematically tractable

Temporal Family of Classifiers (TFC)

$$p(c_k|\mathbf{x}_0, ..., \mathbf{x}_{N-1}) = p(c_k|\mathbf{x}_1) \prod_{2 \le n \le N-1} \frac{p(c_k|\mathbf{x}_n \mathbf{x}_{n-1})}{p(c_k|\mathbf{x}_{n-1})}$$

HMM

Markov Chain Model

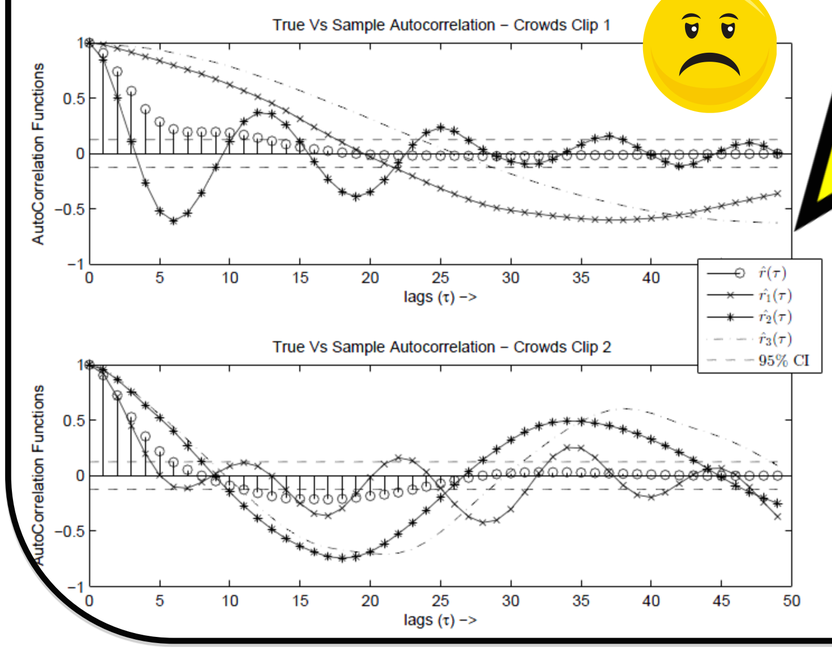
CRF RNN

Large training data, high complexity

Relaxed Assumption - WSS

Failure of Autocorrelation Function

WSS is Covariance-Ergodic in M.S. sense if autocorrelation is absolutely summable (sufficient condition)



$\Delta\chi\Delta\rho \geq \frac{\hbar}{2}$

Ensemble average asymptotically reaches delta function implying i.i.d. case. Relaxed -> WSS

Proposed Test

Signal i

MP based Gabor Representation

VAR on Gabor statistics (P=0 Poles)

VAR based on co-variance statistic

Ljung-Box (LB) portmanteau test on error

No

P=P+1

Yes

P_{opt_1} ••• P_{opt_i}

Average Statistic and 95% CI upper limit < Threshold

Yes

SFC

No

TFC

Table 1. Simulation Results for 45 classes

Class	$\hat{P}_{opt}$	95% UL	Decision	Class	$\hat{P}_{opt}$	95% UL	Decision	Class	$\hat{P}_{opt}$	95% UL	Decision
Airplane	0	1	SFC	Dog	1	2	SFC	Rain	0	0	SFC
Announ.	1	1	SFC	Elevator	1	1	SFC	Restaurant	1	1	SFC
Applause	1	1	SFC	Fan	0	1	SFC	Sheep	2	3	TFC
Baby	2	2	SFC	Footsteps	1	1	SFC	Ship	3	4	TFC
Bell	3	3	TFC	Gun	1	2	SFC	Snore	1	2	SFC
Bird	1	1	SFC	Helicopter	3.5	4	TFC	Stream	0	0	SFC
Boat	1	1	SFC	Horse	1.5	3	TFC	Table tennis	2	3	TFC
Bus	0	1	SFC	Insect	1	1	SFC	Tank	1	2.5	TFC
Car	2	2	SFC	Interior	0	1	SFC	Telephone	0	1	SFC
Chicken	1	1	SFC	Keyboard	1	1	SFC	Thunder	1	1	SFC
Clock	2	6	TFC	Market	1	1	SFC	Traffic	1	1	SFC
Cow	3	4	TFC	Motorcycle	1	2	SFC	Train	1	1	SFC
Cricket	3	4	TFC	Museum	0	0	SFC	Vacuum	1.5	2	SFC
Crowds	1	1	SFC	Pig	1	2	SFC	Water	1	1	SFC
Cutlery	1	2	SFC	Playground	2	2	SFC	Wind	1	1	SFC

Discussion &Future Work

- Out of 45 classes, 10 need temporal modeling. These sounds have temporal evolution and hence generally modeled using left-right HMM.
- Explore generative modeling of unstrucutred sounds in future work.

chachada@usc.edu

Ming Hsieh Institute

Ming Hsieh Department of Electrical Engineering