Graph Theory for the Discovery of Non-Parametric Audio Objects

Christopher Srinivasa*, Martin Bouchard*, Ramin Pichevar^, and Hossein Najaf-Zadeh^

University of Ottawa*

Communications Research Centre Canada^

Outline

I. Problem Statement
II. Pre-processing
III. Proposed Object Extraction Framework
IV. Performance
V. Summary and Future Work

Section I. Problem Statement

Section I: Audio Coding

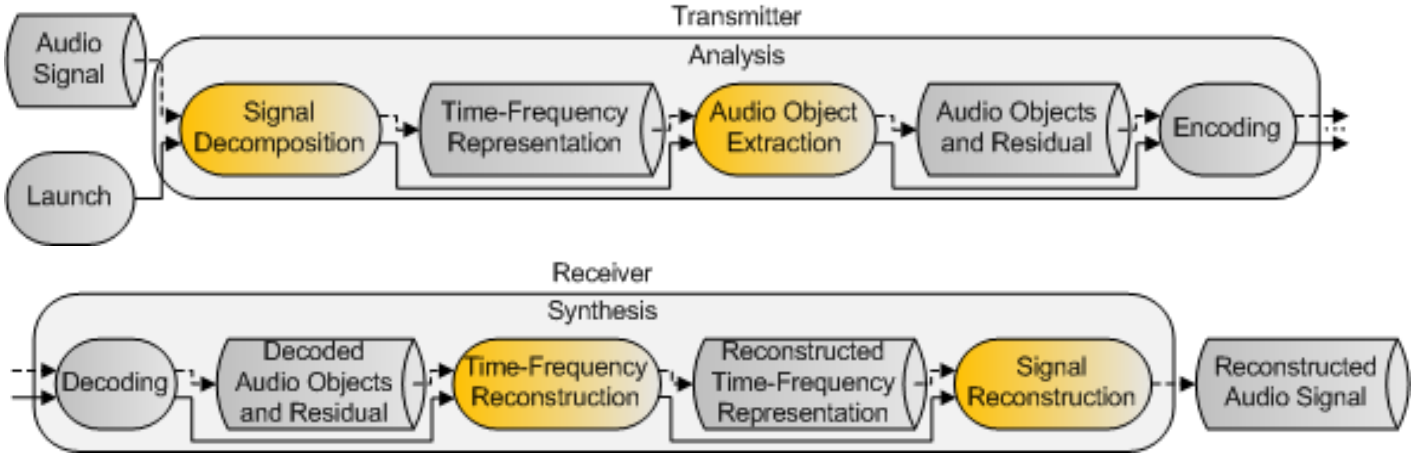
- Current object-based audio coding methods are parametric: describe a signal as a set of concise audio objects using prior knowledge about the structure of audio
- Example is the harmonic sound: combination of sinusoids described by single parameter (i.e. the fundamental frequency)
- Set of parameters available to model signal must be defined a priori
- Choice of parameters available may not always be optimal

Section I: Proposed Method

- Proposed method defines new type of objects: Non-Parametric Objects (NPO)
- NPOs: combinations of coefficients occurring more than once in time-frequency representation
- Allows extraction of broader class of audio objects (no predefined parameterization space imposed on extracted objects)
- Object: sound which minimizes external cost function unrelated to shape

Section I: Proposed Method

- Proposed object extraction framework if used in full audio coder

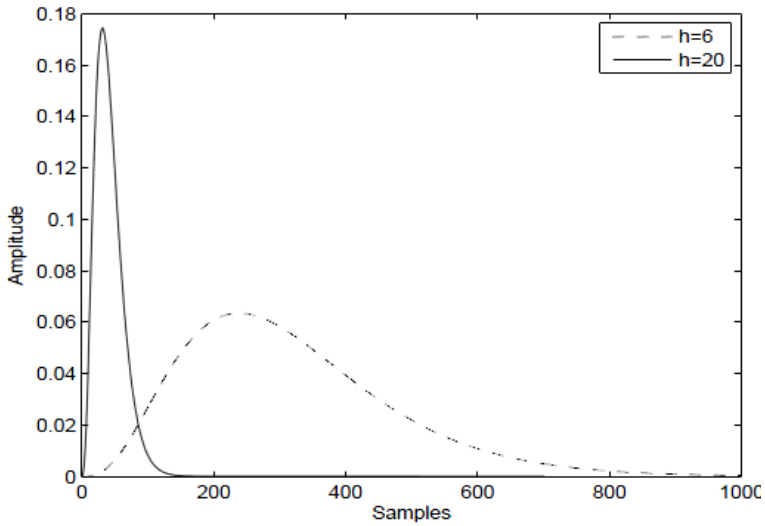


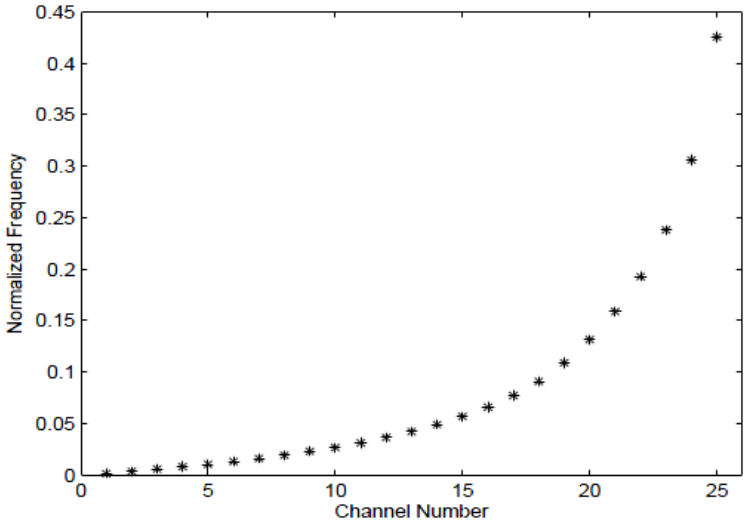
- Focus: audio object extraction framework, pre-processing and reconstruction modules

Section II. Pre-processing

Section II: Signal Decomposition

- Ear deciphers signal based on its frequency content via hair cells on membrane
- Achieved by evaluating signal against set of kernels representing different frequencies
  - Algorithm: Perceptual Matching Pursuit
  - Kernels: Gammatones



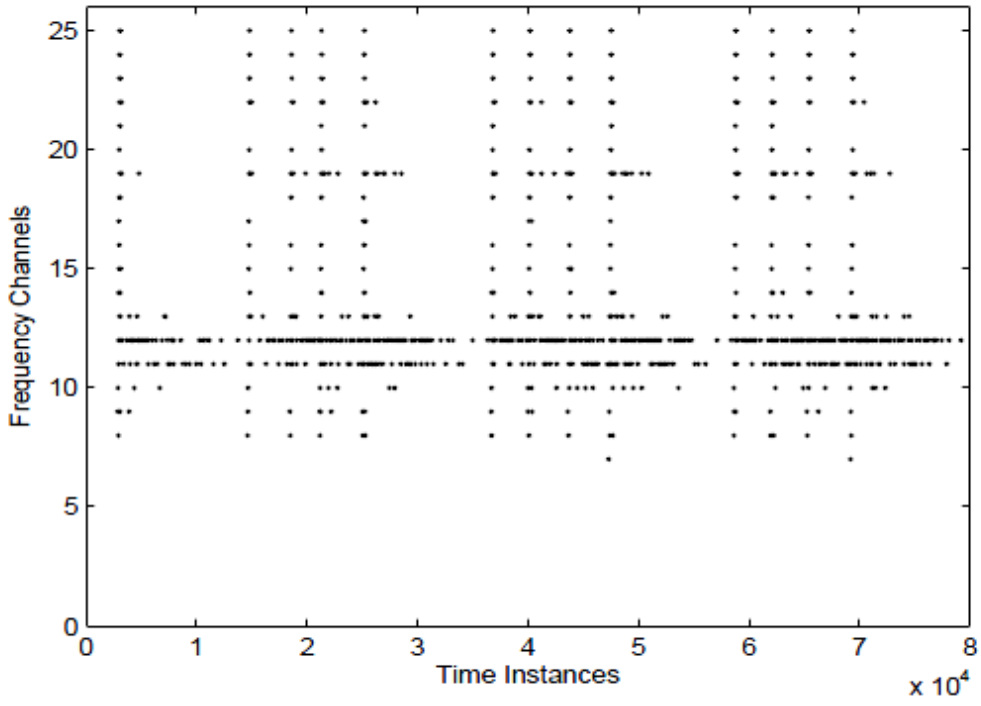


Section II: Representation

- Resulting representation is a spikegram
- Each spike described by four components

$$\mathbf{s}^i = \left\langle s^i_{mag}, s^i_{ang}, s^i_{pos}, s^i_{chan} \right\rangle$$

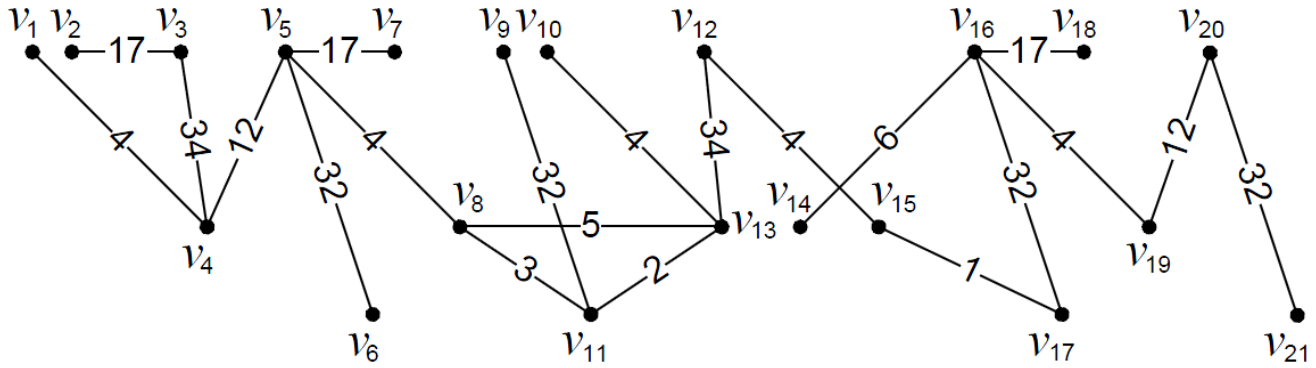
- Signal reconstructed by summing all spikes



Section III. Proposed Object Extraction Framework

Section III: Audio Object Extraction

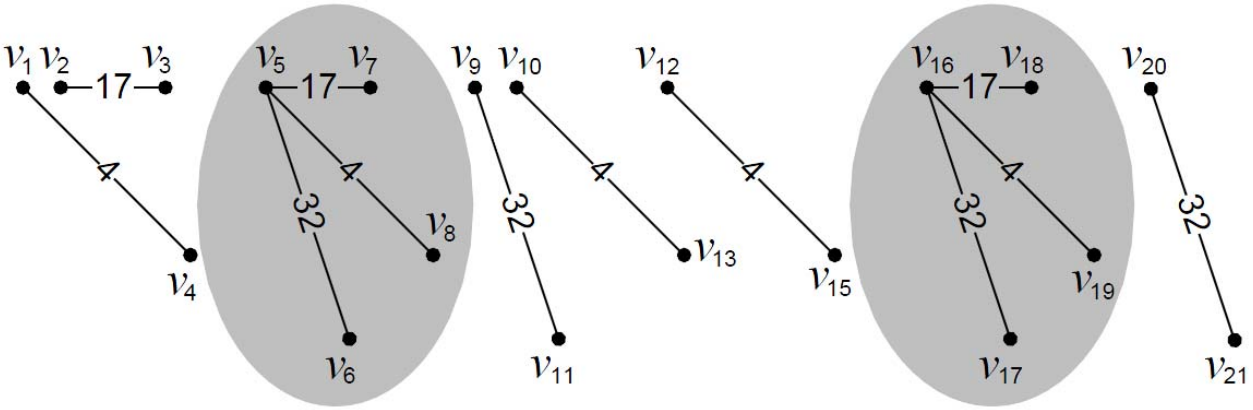
- Each spike a vertex, linked to other vertices via labelled edges based on similarities of vertices and their relationships



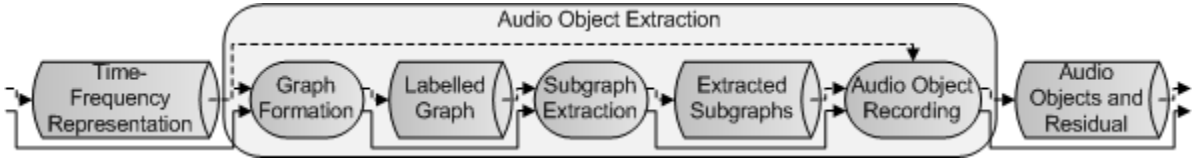
- Objects mined as frequent subgraphs where each subgraph is an instance of an object

Section III: Audio Object Extraction

- Each object characterized by the recurring edge labels involved in each of its instances
- For object recording intent, only stars are valid



- Framework summarized as follows



Section III: Graph Formation

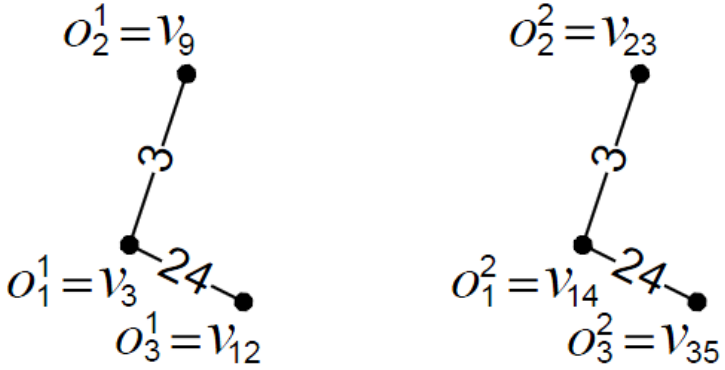
- Can form graph with any spike components
- As example, use only frequency and time
- Each edge described by a feature vector

$$\mathbf{z}^{e_{(i,j)}} = \left\langle s_{chan}^{v_i}, s_{chan}^{v_j}, s_{pos}^{v_j} - s_{pos}^{v_i} \right\rangle$$

- Edges are labelled by clustering their vectors
- Quality Threshold clustering ensures all vectors assigned to same cluster do not deviate from each other by more than predefined threshold(s)

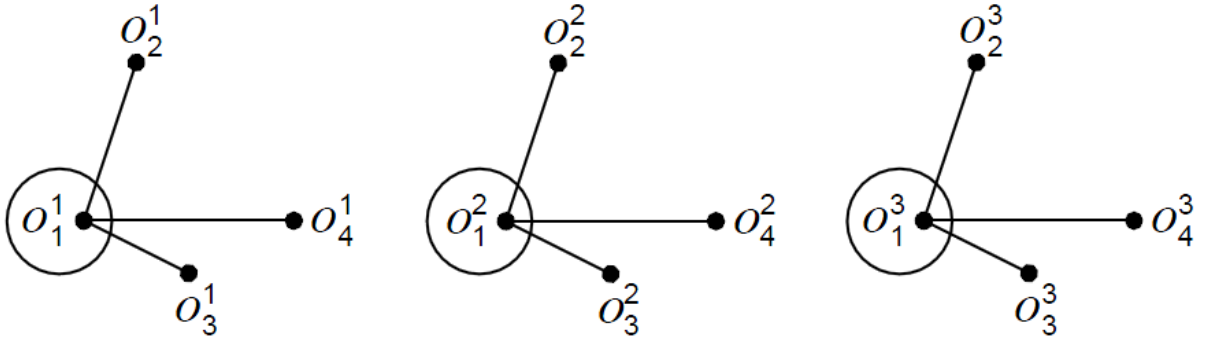
Section III: Subgraph Extraction

- Frequent stars mined with iterative star extraction algorithm and cost function
- Cost function based on fixed-length bit scheme
- Star search finds all frequent stars in graph and computes extraction cost for each one
- Frequent star with minimal cost is recognized and its instances are recorded and removed
- Each minimal frequent star becomes an object



Section III: Audio Object Recording

- Each object is recorded with the anchor point in each instance and other spikes recorded as representative relationships



- In the example, only the time and frequency spike components involved in the objects are recorded
- Objects can be reconstructed by placing copies of the representative relationships at each instance

Section IV. Performance

Section IV: Setup

- Tested on five audio excerpts
- Gain evaluated using a cumulative cost function
- All spike components considered for overall gain
- Relative gain factors in only the compressed spike components (i.e. time and frequency)
- Quality evaluated using Signal to Noise Ratio, Segmental Signal to Noise Ratio, and Perceptual Evaluation of Audio Quality model

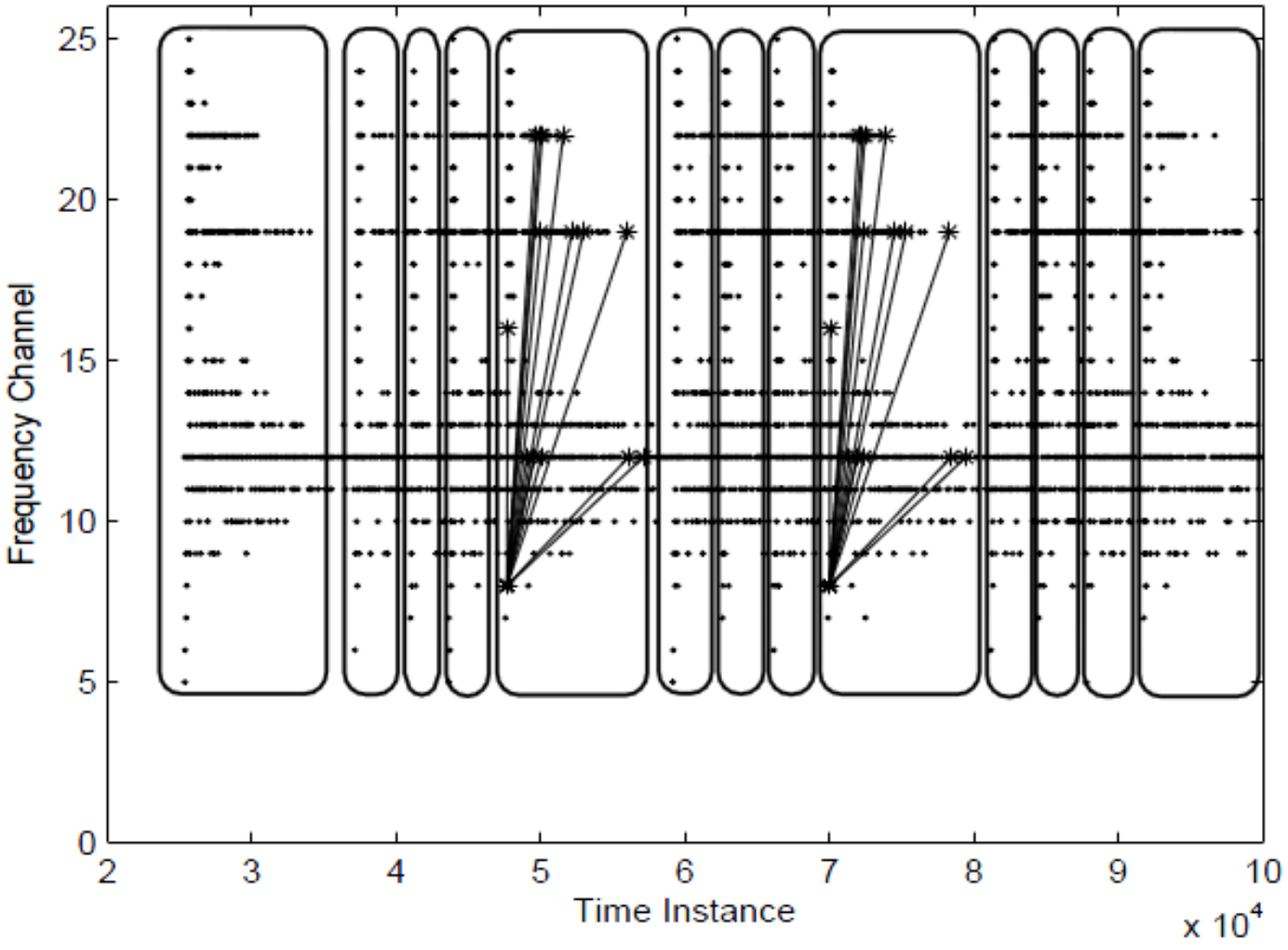
Section IV: Results

- Results show average overall and relative gains of 15.90% and 23.53% with a PEAQ score of -0.395

Audio	Objects	Gain (%)		SNR (dB)	SSNR (dB)	PEAQ Score
Castanet	1178	O	28.67	10.41	12.18	-0.752
		R	43.00			
Vibra-phone	2305	O	24.44	21.18	16.06	-0.734
		R	36.14			
Female Speech	1318	O	8.71	16.16	17.96	-0.191
		R	12.71			
Male Speech	1061	O	8.39	19.47	20.45	-0.125
		R	12.24			
Piano	1395	O	9.28	20.60	21.03	-0.175
		R	13.54			

Section IV: Results

- New types of objects also found



Section V. Summary and Future Work

Section V: Summary

- Novel graph theoretic framework applied to discover new types of audio objects: NPOs
- Shape of NPOs not restricted by any a priori psychoacoustic knowledge
- New types of objects discovered while achieving compression and maintaining a high audio quality

Section V: Future Work

- Further evaluate performance with informal/formal listening tests and larger corpus
- Create a full end-to-end audio coder centered around the proposed framework
- Make quantitative comparisons with other coders in industry

Questions