

# Biologically-Inspired Sparse Coding of Music

Ramin Pichevar, Hossein Najaf-Zadeh, and Hassan Lahdili  
Communications Research Centre, Ottawa, Canada  
emails: {Ramin.Pishevvar, Hossein.Najafzadeh, [Hassan.Lahdili](mailto:Hassan.Lahdili@crc.ca)}@crc.ca

## **Abstract**

*This research deals with the generation of auditory-inspired spectro-temporal features aimed at audio coding. To do so, we first generate sparse audio representations we call spikegrams, using projections on gammatone/gammachirp kernels that generate neural spikes. Unlike Fourier-based representations, these representations are powerful at identifying auditory events, such as onsets, offsets, transients and harmonic structures. We show that the introduction of adaptiveness in the selection of gammachirp kernels enhances the compression rate compared to the case where the kernels are non-adaptive. We also integrate a masking model that helps reduce bitrate without loss of perceptible audio quality. We then quantize coding values using the genetic algorithm that is more optimal than uniform quantization for this framework. We finally propose a method to extract frequent auditory objects (patterns) in the aforementioned sparse representations. The extracted frequency-domain patterns (auditory objects) help us address spikes (auditory events) collectively rather than individually. When audio compression is needed, the different patterns are stored in a small codebook that can be used to efficiently encode audio materials in a lossless way. The approach is applied to different audio signals and results are discussed and compared. This work is a first step towards the design of a high-quality auditory-inspired “object-based” audio coder that can be used in the music technology/industry.*

## **Introduction**

Non-stationary and time-relative structures such as transients, timing relations among acoustic events, and harmonic periodicities provide important cues for different types of audio processing techniques including audio coding, speech recognition, audio localization, and auditory scene analysis. Obtaining these cues is a difficult task. The most important reason why it is so difficult is that most approaches to signal representation/analysis are block-based, i.e. the signal is processed piecewise in a series of discrete blocks. Therefore, transients and non-stationary periodicities in the signal can be temporally smeared across blocks. Moreover, large changes in the representation of an acoustic event can occur depending on the arbitrary alignment of the processing blocks with events in the signal. Signal analysis techniques such as windowing or the choice of the transform can reduce these effects, but it would be preferable if the representation was insensitive to signal shifts. Shift-invariance alone, however, is not a sufficient constraint on designing a general sound processing algorithm. A desirable representation should capture the underlying 2D-time-frequency structures, so that they are more directly observable and well represented at low bit rates [1]. These structures must be easily extractable as auditory objects for further processing in coding, recognition, etc.

The aim of our work is to propose an auditory-inspired coding scheme, which includes many characteristics of the auditory pathway such as sparse coding, masking, auditory object extraction, and recognition.

## **Proposed Approach**

We propose a new biologically-inspired paradigm for universal audio coding based on neural spikes. Our proposed approach is based on the generation, by matching pursuit, of auditory-inspired sparse 2-D representations of audio signals, dubbed as spikegrams. We show that the use of an adaptive scheme

that dynamically changes the tuning parameters in the gammachirp kernels considerably decreases spike counts and the associated bitrate to code audio materials for a given quality when compared to the non-adaptive paradigm. We then apply simultaneous, temporal backward, and temporal forward masking models to further reduce spike counts for a given perceived quality and from there propose the Perceptual Matching Pursuit [2] that defines an automatic stop criterion based on the masking threshold. We then derive the optimal amplitude quantization for our spikegrams using genetic algorithm (GA) and proposed a much faster suboptimal piecewise linear quantization scheme with audio quality close to the optimal solution obtained by GA [3].

We finally propose a fast (faster than the MP stage) frequency-domain audio object (episode) extraction algorithm based on the generation of spikegrams. The advantage of such an algorithm stems in the fact that spikegrams are representations of discrete events that can be mined easily by known approaches. This is in contrast with raw or frequency-domain representations of signals (i.e., spectrogram) in which each sample can take so many values and where data mining is difficult to perform. We then applied our proposed technique to audio coding and obtained promising results for the lossless coding of frequency-based information. Informal listening tests show that the overall system gives high quality (scores above 4 on the ITU-R 5-grade impairment scale) and has the potential to achieve the target 44.1 kbps [4].

### **Future Work**

In a future work, we will extract the structural dependencies of spike amplitudes and/or other parameters in the spikegram such as the chirp factor, etc. We will also investigate the design of a generative neural model based on spikegrams. Formal subjective listening tests for the overall system will be conducted later on. In order to speed up the spikegram extraction of audio signals, we have conducted preliminary tests on replacing the Matching Pursuit stage by neural circuitry that can be implemented on embedded and parallel hardware [5]. We will further explore this avenue in a future work. The application of different ideas outlined in this article (i.e., pattern recognition and masking model) are not limited to spikegrams and can be applied to other sparse representations in the literature. In addition, the frequency episode discovery algorithm discussed in this article can be used in speech recognition, sound source separation, audio classification, etc. in future work.

[1] E. Smith and M. S. Lewicki, Efficient Auditory Coding, *Nature*, 439 (7079), 2006.

[2] H. Najaf-Zadeh, R. Pichevar, L. Thibault, H. Lahdili, "Perceptual Matching Pursuit", Audio Engineering Society Conv., Amsterdam, The Netherlands.

[3] R. Pichevar, H. Najaf-Zadeh, L. Thibault, H. Lahdili, "Entropy-Constrained Spike Modulus Quantization In a Bio-inspired Universal Audio Coder ", European Signal Proc. Conf., 2008, Lausanne, Switzerland.

[4] R. Pichevar, H. Najaf-Zadeh, and L. Thibault, "A Biologically-Inspired Low-Bit-Rate Universal Audio Coder", Audio Engineering Society Convention, Vienna, Austria, 2007

[5] C.J. Rozell, D.H Johnson, R.G. Baraniuk, and B.A. Olshausen. Sparse coding via thresholding and local competition in neural circuits. *Neural Computation* 20, 2526-2563, 2008.