FODAVA-Partner: Visualizing Audio for Anomaly Detection

This proposal addresses the type of audio anomalies that human data analysts hear instantly: angry shouting, trucks at midnight on a residential street, gunshots. The human ear detects anomalies of this type rapidly and with high accuracy; for example, rifle magazine insertion clicks are detected with 100% accuracy at 0 dB SNR in white noise, babble, or jungle noise. Unfortunately, a data analyst can listen to only one sound at a time. Visualization shows analysts many sounds at once, possibly letting them detect anomalous sounds several orders of magnitude faster than "real time." This proposal aims to to render large audio data sets, comprising thousands of microphones or thousands of minutes, in the form of interactive graphics that reveal important anomalies at a glance.

Intellectual Merit: Data transformation for visual analytics is like a lossless encoding of the data: the optimal data transformation should highlight apparently anomalous events, but should also let the human analyst perfectly reconstruct even the most apparently mundane places and times. For this purpose we propose a two-stream data transformation. In the first stream, raw audio features are optimized to let the analyst understand the audio's semantic content in every portion of a large database. We propose specifically that audio representations must include: spectrograms or wavelet representations, smoothly zoomable over a wide range of temporal and frequency resolutions; correlograms and other measures of multi-periodic harmonicity; and summary measures such as spectral center of gravity and zero-crossing rate. In the second stream, model-based features are optimized to highlight space-time coordinates or time-frequency coordinates that are "anomalous" in the technical information-theoretic sense: events whose feature vectors fail to match the statistics of the dataset, according to one of several available statistical models. Audio event detection is not yet accurate enough to be trusted without human intervention: instead, we propose to train and test complementary best-of-breed anomaly detection measures, whose conflicting claims can be appropriately weighed by an experienced human analyst.

Broader Impact: All software and datasets constructed for this research will be published on the web, using open-source/open-data licenses comparable to those governing redistribution of our Syzygy visualization toolkit and AVICAR audiovisual speech corpus. All experimental and theoretical results will be presented at FODAVA project meetings, and in professional conferences and journals.

Data transformation algorithms developed in this research will be tested in two visualization testbeds, both of which will be published (data and software) on the web: the Multi-Day Audio testbed, and the Milliphone. First, the Multi-Day Audio testbed will ask human subjects, using a zoomable spectrogram-like interface with novel data transformations, to rapidly find anomalous events (either natural anomalies, e.g., airplane fly-overs, or anomalies planted by the experimenters) in a multi-day audio recording. The target application for the Multi-Day Audio testbed will be a portable audio browser for first responders. The Milliphone (1000 microphones) will summarize the recordings acquired by 1000 microphones distributed in public locations throughout an urban landscape. The target application for the Milliphone will be a three-dimensional display, in a data analysis center or situation room (simulated with the Beckman CUBE virtual reality theater). In the Milliphone display, each microphone signal will be displayed as a brightly colored thread hanging vertically above the city map, with height above the map depicting time (before present). Color and brightness of each thread, at each time, convey summary audio and model-based features including the probability and type of anomaly. Both testbed applications will be used to quantify the gains achieved by proposed methods of data transformation: a successful outcome will be a visual summary that lets the analyst detect most anomalies immediately (about 10,000× faster than real time), and all anomalies after brief interactive exploration (about 1,000× faster than real time).