Andrew and Erna Viterbi
Faculty of Electrical Engineering
Electronics
Computers
Communications

SIPL
Signal and Image Processing Lab

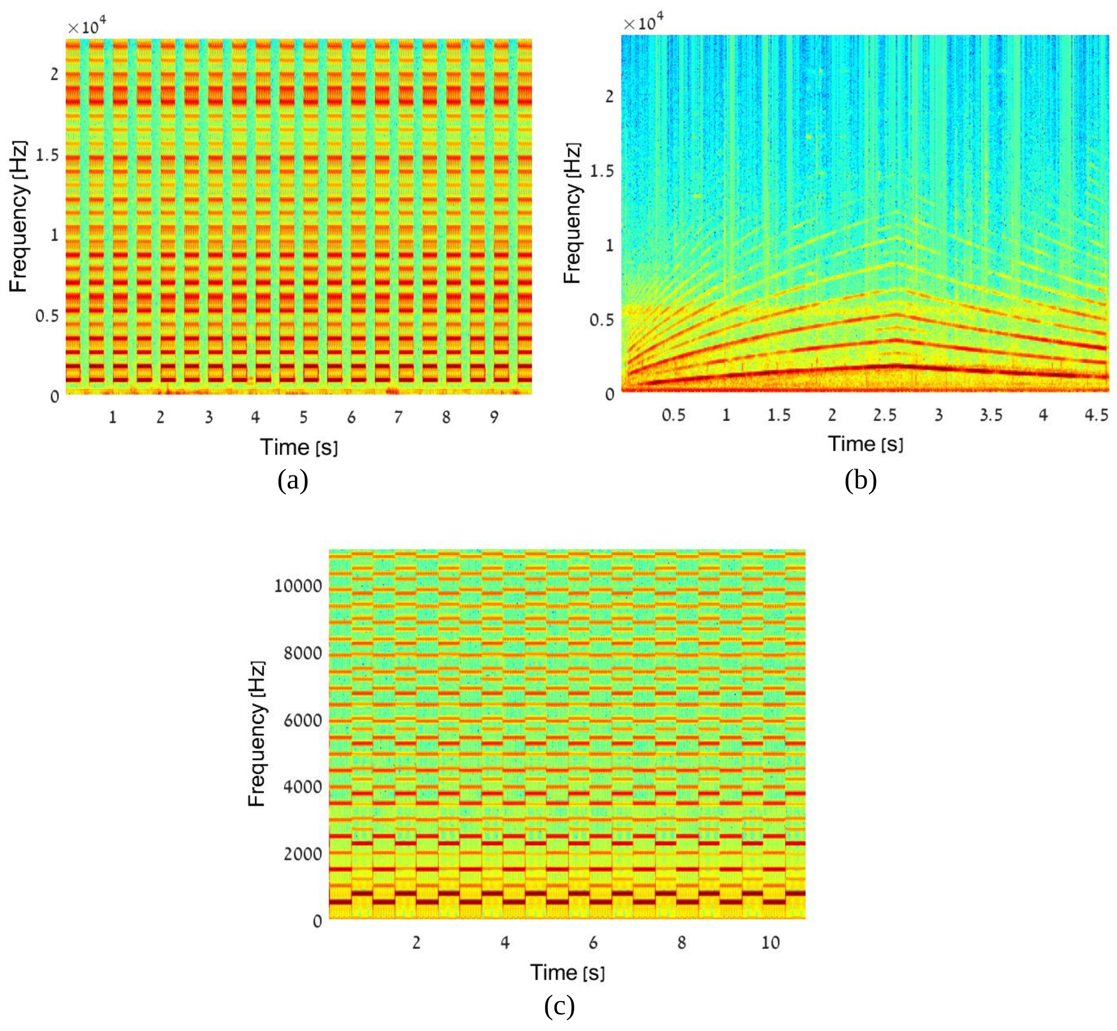
TECHNION
Israel Institute of Technology

Detection of Alarm Sounds in Noisy Environments

Dean Carmel, Ariel Yeshurun, Yair Moshe

Introduction

- Alarm sounds play an important role in everyday life since they warn people of hazardous situations.
- Automatic detection of this class of sounds can contribute to the independence and safety of hearing impaired or distracted people.



Spectrograms of 3 types of alarm sounds: (a) Alarm clock: Pulsed alarm - Repetition of the same sound with a silent gap between each instance. (b) Ambulance driving away: Siren - The frequency continuously changes. (c) Fire alarm: Alternating alarm - Two different alternating tones with no silence gap between them

Previous Works

- Several previous works tried to deal specifically with the task of alarm sound detection.
- Most works try to detect only particular alarms, usually sirens of emergency vehicles of a specific country.
- Many of these works do not perform well out of laboratory conditions since they do not model well enough ambient background.

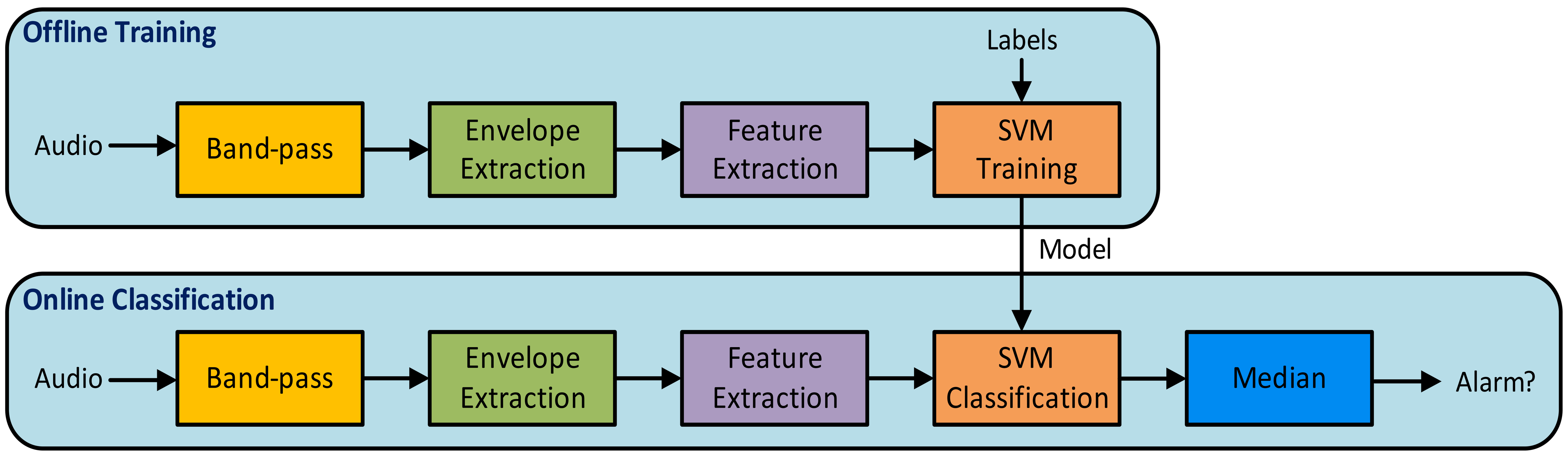
Goals

- A **generic** technique for the detection of alarm sounds in noisy environment.
  - **Not limited to particular alarms** and can detect most electronically generated alerting sounds.
- Detect alarm sounds with a short time delay.

Challenges

- Distinctive characteristics of alarm sounds are not formally defined.
- Lack of standardization of alarm sounds between different countries and organizations.
- Highly noisy environments (traffic etc.)
- Doppler Effect.

Proposed Technique



In the offline training phase, audio frames labeled as alarm or no-alarm are used to train an SVM classifier. This trained classifier is used in the online classification phase to classify new unlabeled audio frames as alarms or no-alarms.

Feature Extraction

1. **Time domain features** - pitch frequency, short time energy, zero crossing rate.
2. **Frequently domain features** - MFCC, spectral flux, spectral roll-off, spectral centroid, spectral flatness.
3. **Wavelet-based features** - capture time and frequency localized information and includes coefficients of the discrete wavelet transform (DTW), and wavelet packet transform (WPT).

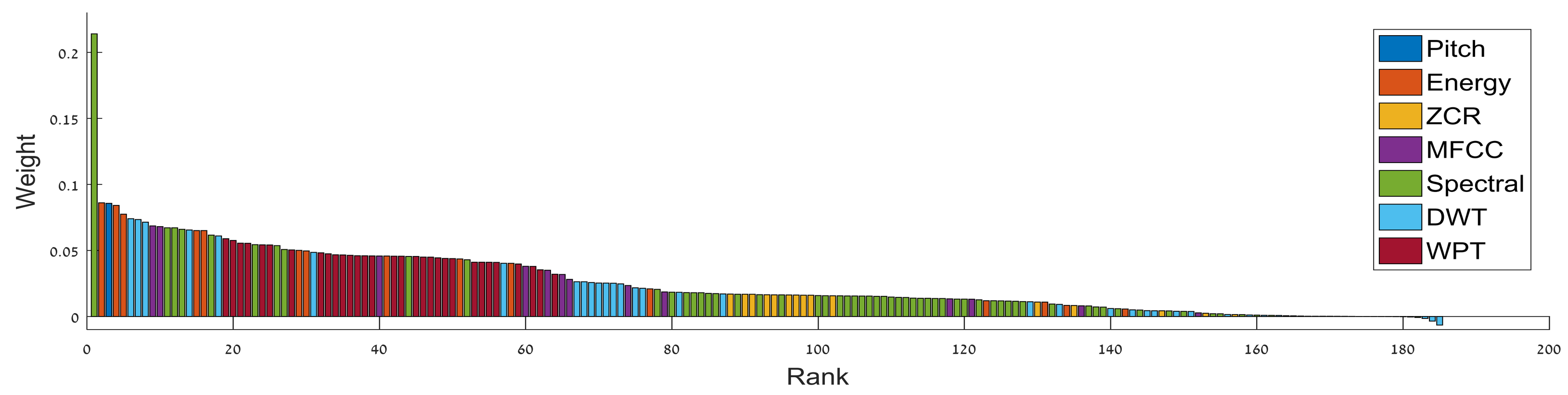
For each feature, we compute 18 statistics of overlapping sub-frames:

- maximum (max)
- minimum (min)
- mean
- median
- standard deviation (std)
- max/mean
- max/median
- $\text{std}^2/\text{mean}^2$
- 2^{nd} smallest value
- 3^{rd} smallest value
- 4^{th} smallest value
- 5^{th} smallest value
- mean of 4 smallest values
- mean of values $> 10^{-6}$
- $\frac{\#\text{values} > \text{mean}}{\#\text{values}}$
- $\frac{\#\text{values} > 0.1\text{mean}}{\#\text{values}}$
- $\frac{\#\text{values} > 5\text{mean}}{\#\text{values}}$
- $\frac{\#\text{values} > 0.4}{\#\text{values}}$

This results in a feature vector of size 200.

Feature Selection

- We use the ReliefF algorithm [Kononenko, et al., 1997].
  - Uses k-NN to decide the importance of each feature.
- Features that do not contribute to a higher classification accuracy are removed and thus computational and storage complexity are reduced.



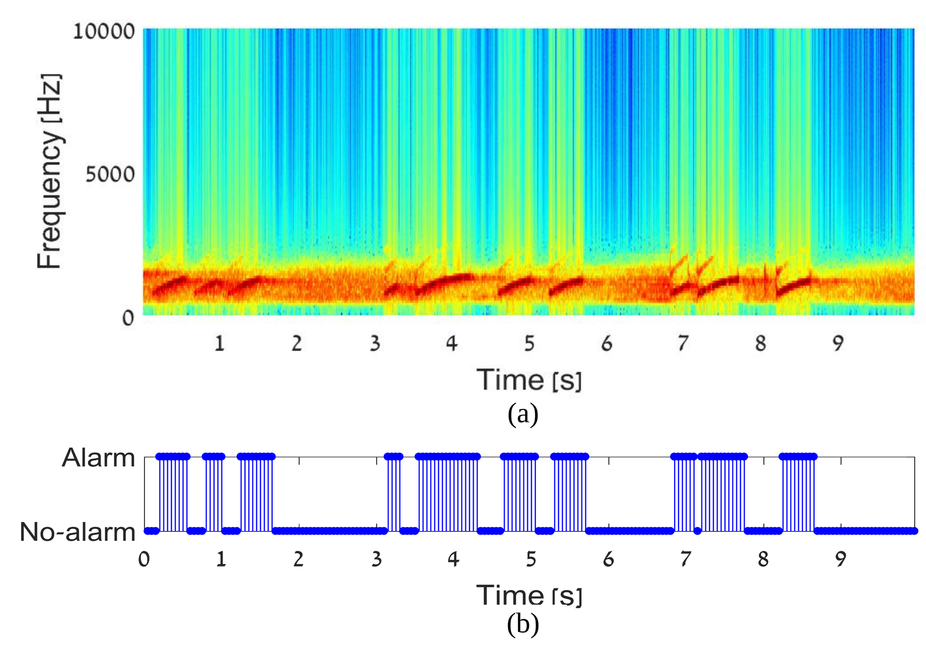
Feature ranks and weights computed by the ReliefF algorithm. Weight ≤ 0 indicates features that do not contribute to a higher classification accuracy and are therefore removed.

Database

The database contains 70 audio signals, 35 alarm sounds and 35 noises. Sounds were collected by searching the web and by making recordings around the home and office.

Results

- ✓ 98% accuracy per audio frame.
- ✓ 0.4% false positive rate.
- ✓ Good recognition of low SNR alarm sounds.



(a) A spectrogram of a police car operating its siren in intermittent bursts and (b) alarm/no-alarm classification results of its audio frames. Classification results are 100% accurate although the siren operates in short bursts and is interleaved with noise.

Conclusions

- Successful detection of various types of alarm sounds.
- Short detection delay (200 ms).
- Robustness to noise and other real-life conditions.

Partly supported by החממה החברתית טכניון social hub Technion

EUSIPCO2017