AUDITEUR : A Mobile-Cloud Service Platform for Acoustic Event Detection on Smartphones

Shahriar Nirjon, Robert F Dickerson, Philip Asare, Qiang Li, Dezhi Hong, John A Stanovic, Pan Hu, Guobin Shen, Xiaofan Jiang

PRESENTER : MADHUMITHA SUBRAMANIAN
INTRODUCTION

- **Auditeur (Listener) – Acoustic Event Detection Platform for smartphones.**
- **Backed by a cloud service- to store sound clips (user contributed) and generates classification plan.**
- **Energy efficient/Context aware.**
- **Increases device life time.**
- **Examples of acoustic events : Music recognition, speaker identification, emotion/stress detection etc.**
- **Developer platform (not just a detection system).**
MOTIVATION

- For specialized tasks, smartphones make use of web services.
- Example: Maps, Play, Gmail by Google; Translator by MS Hawaii
- Existing acoustic event detectors: JigSaw, Soundsense

ADVANTAGES OF AUDITEUR

- API is simple and easy to learn
- Supports generalization and personalization
- Efficient in energy consumption, computation
- Uninterrupted internet connection is not required
- Adaptive sound recognition
CONCEPT

- Used “Tagged Soundlets” for classification purpose.
- Tagged Soundlets – short duration audio clips with user given tags.
- Cloud’s purpose – Host tagged soundlets, provide services, upload new soundlets, obtain classification plan.
- Classification plan – Contains acoustic processing pipeline configurations.
- It is a hybrid of in-phone sound recognizer and cloud assisted sound recognizer.
- Signal processing, feature classification performed in phone.
- Storing sound clips, building new classifiers, creating plans are done in cloud.
For Developers

To create app that classifies music according to genres and suggests same genre music.

For Researchers

Study on social interaction.

Voice, laughter, yelling sounds stored and classified.

End Users

Example: A mother with a new born child.

If the mother is working in some other room, she can custom the app to notify her when particular sound occurs in the baby’s room.
TAGGED SOUNDLETS

- Contains audio clips + user given tags + phone generated context.

- Tags: Container tags and Content tags.
  - Content tag: Describes sound. Ex: { voice, female, Alice }
  - Look-for tag: Soundlets to be classified. Ex: { voice, Alice }
  - Within tag: Sounds in environment. Ex: { voice, printer, phone }
  - Container tag: Describes background. Ex: { office }
Phone Contexts: Additional information about soundlet.

Three types of Contexts

Phone Location

Position of the phone with respect to the body

Environmental noise level
AUDITEUR SYSTEM

IN PHONE PROCESSING

- API to record, add tags and upload soundlets to the cloud
- Energy consuming process but done only once.
- Acoustic Event Detection
IN CLOUD PROCESSING

- Energy aware acoustic detection plan generated upon request.
- Training set created.
- Many classifiers are trained and the best one is chosen.
KEY FEATURES

- People in the Loop
- Personalization and Generalization
  - Public Space – for developers
  - Private Space – ensures privacy
- Cloud-Directed on Device Processing
  - Configuration file for every problem
- Context Awareness
- Efficiency
IN-PHONE COMPONENTS

- Sound Engine performs the actual communication task.
- Acoustic Processing Units (APUs) form the acoustic processing pipeline.

ACOUSTIC PROCESSING PIPELINE

- Preprocessing: captures audio, converts byte stream into frames.
- Extract frame features.
- Decide whether or not a frame should be processed.
- Window Formation

*Figure 6: The acoustic processing pipeline in Auditeur.*
IN-PHONE COMPONENTS

- PHONE CONTEXT GENERATION
  - Location based-GPS (Indoors/Outdoors)
  - Context checking can be enabled/disabled.

- COMMUNICATIONS TO THE CLOUD
  - Web HTTP interface to upload audio.
  - Amazon EC2 cloud.
  - Soundlets are serialized into JSON objects.
IN CLOUD COMPONENTS- CLASSIFICATION PLAN

Main goal: To generate energy aware event detection plan

Training set generation

Choose subset of soundlets

Environment specific training- container tagged soundlets chosen

Soundlets with content tags marked positive

Acoustic features extracted.
Feature selection algorithm

Minimizing energy consumption is the goal - Feature subset is taken

A feature is good if it is relevant and non redundant

Calculate entropy of a feature.

When another feature is added, find information gain

To find optimal subset of features, recurrence relation is formed and solved.

\[ f_j(b) = \begin{cases} f_{j-1}(b) & \text{if } X_j \text{ is not taken} \\ f_{j-1}(b - e_j) \cup \{X_j\} & \text{otherwise} \end{cases} \]
IN CLOUD COMPONENTS

TAG MATCHING

- Used for admission control and to tag untagged soundlets
- Distance based outlier detection technique
- Computation of similarity scores of feature vector against existing soundlets
- Return the top k tags that are similar
STORAGE AND CACHING

- MongoDB, a NOSQL database is chosen
- When classification plan is requested, local cache is first searched
- If not found, new plan is generated and stored in cache
- Cache is updated to avoid stale data
- On demand creation of classifiers
EVALUATION

METRICS

- CPU and Memory consumption
- Energy consumed for different kinds of sounds
- Accuracy-Energy tradeoff

DATA COLLECTION

- First source – People of Virginia and Beijing
- Second source - Internet
MEASUREMENTS

CPU AND MEMORY

<table>
<thead>
<tr>
<th></th>
<th>CPU</th>
<th>Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditeur (silence)</td>
<td>6%</td>
<td>6.5 MB</td>
</tr>
<tr>
<td>Auditeur (active)</td>
<td>16%</td>
<td>11.8 MB (22 MB)</td>
</tr>
</tbody>
</table>

ENERGY

- Kill all background services, set brightness to minimum.
- Two min long audio file.
- Feature selection takes up 98% of totally consumed energy.

Figure 9: Feature extraction accounts for 98.48% of the total energy consumption of the processing pipeline.
ENERGY-ACCURACY TRADEOFF

- Trades off less informative features for longer lifetime.

- EVALUATING FEATURE SELECTION
  - Comparison with WEKA
  - Auditeur considers energy bound
ENERGY EFFICIENCY

Run auditeur in two android phones—one with energy bound/one without

PROCESSING DELAY

Auditeur is sensitive to energy bound

Processing delay decreases as lifetime increases

Overall delay is always less than 1s for 1s frames
THANK YOU