Audio Event Recognition: Pathways to Impact

Sacha Krstulovic
V.P. of Technology
Audio Analytic Ltd.
www.audioanalytic.com
Audio Analytic

• AA is commercialising Automatic Environmental Sound Recognition (AESR) for the Smart Home
  ▪ Non-speech, non-music
  ▪ a.k.a. Audio Events Detection (AED)
Pathways to Impact

1. Data → Research → Lab Evaluation → Field Evaluation

Impact (Application)

Informing research topics and methodology?
Pathways to Impact

Data → Research → Lab Evaluation → Field Evaluation

Informing research topics and methodology

Impact (Application)
Market and applications

• AA’s primary target: Smart Home market
  ▪ Supported by distributors.

• Application: Acoustic Ambient Artificial Intelligence
  ▪ “Your home listens for audio events and alerts you or takes appropriate actions”.
  ▪ “Peace of mind.”

• Other markets and applications?
Pathways to Impact

Data → Research → Lab Evaluation → Field Evaluation → Impact (Application)

Informing research topics and methodology
Data

- Smart Home: Indoor sounds!
  - Need more indoors public data sets.
  - Doesn’t reduce the generality of the AESR problem: the Taxonomy of sounds is still generic.
  - May help focusing the research a bit more.
Taxonomy of sounds

Time

Frequency

kHz

kHz
Labelling

• Labour intensive, very costly.
• But it must be done.
• ... And it must be done well!
• Cost reduction strategies?

It’s a FROG!
Labelling

• “Bucketing” approach: labelling the contents of “coarse” chunks.

[Foster & al, WASPAA 2015 – DCASE task 4]
Fast, but data is “impure”.

• Semi-supervised and unsupervised approaches are possible. But still require some human checking and/or hand correction.
Data collection
Pathways to Impact

Data → Research → Lab Evaluation → Field Evaluation → Impact (Application)

Informing research topics and methodology
Robustness

• Same sound captured by various consumer products

• Clearly audible channel differences!

• => Research topic: Robustness!
Meta-parameters

• Optimise:
  ▪ Number of DNN layers
  ▪ Number of Gaussian clusters
  ▪ Feature set
  ▪ Learning rates
  ▪ Etc.

• Optimisation
  => Evaluation metrics?
Pathways to Impact

- Data
- Research
- Lab Evaluation
- Field Evaluation
- Impact (Application)

Informing research topics and methodology
Evaluation

• Detection is traditionally evaluated over closed data sets, AFTER a classification decision has been made:
  
  Precision = TP / (TP + FP)
  Recall = TP / (TP + MD)
  F-score = 2 * P * R / (P + R)

• P and R are OK to compare systems, but P heavily depends on the choice of non-target set! (Data set size and priors.)

• In practical reality, non-targets are very important:
  
  ▪ Open set: a real sound detection system will be continuously exposed to non-target sounds.
  ▪ True Positive units are easy to define (e.g., the extent of a baby cry) but what are the False Positive units?

  Could use blocks/chunks, but pros and cons.
  [Heittola & al, EURASIP J. on Audio, Speech & Music Processing 2013]
Open set non-target

Confusion matrices:

- CLEAR eval 2006
  13 classes
  - 13 misclassifications

- DCASE 2013
  16 classes
  - 16 misclassifications

- 24/7 Sound Event Recognition
  - 1 misclassification
  - 100% accuracy
Decision

- Detection = threshold on scores

- **GMMs** -> likelihood ratio between target model and world model
- **SVMs** -> deviation from the margin
- **DNNs and RNNs** -> single output, class membership probability

- A given choice of threshold defines a single **Operation Point**: compromise between **False Alarm** and **Missed Detection** rates.
Operation point trade-off

- A system can be set to detect sounds more or less conservatively. Thought experiment:

**less sensitive / more conservative**

Less sounds get through -> less FAs but more MDs.

3 True Positives
NO False Alarms.

5 True Negatives:
×××××
2 Missed Detections:
××

**more sensitive / more permissive**

More sounds get through -> less MDs but more FAs.

None of the non-target sounds went through, but some of the target sounds were missed.

5 True Positives
3 False Alarms

2 True Negatives:
××
NO Missed Detections.

None of the target sounds were missed, but some non-target sounds went through.
Application-independent evaluation

• But: the goal is to evaluate the models, NOT the wisdom of choice of threshold.

• **DET curves** (Detection Error Trade-off)
  Plot all possible tradeoffs between FA and MD rates by browsing the threshold.

• **EERs** (Equal Error Rates) locate DET curves along the diagonal.

• Lots of work has been done in the domain of Speaker Recognition

  “An Introduction to Application-Independent Evaluation of Speaker Recognition Systems”
  D. van Leeuwen and N. Brummer, 2007
  Speaker Classification I, Springer
  Vol. 4343 Lecture Notes in Computer Science, pp 330-353
Decision Cost Function

• Customer X: “I want to minimise false alarms to minimise customer support requests.”
• DETs crossing: now which system is best?
• DCF, “Decision Cost Function”

$$DCF = C_{miss} \times P_{miss|target} \times P_{target} + C_{FA} \times P_{FA|nonTarget} \times (1 - P_{target})$$

Involves costs $C_{miss}$ and $C_{FA}$, as well as prior $P_{target}$.

• In speaker recognition, usually (and arbitrarily) $C_{miss} = 10, C_{FA} = 1, P_{target} = 0.01$

But for sound recognition, $P_{target}$ can be infinitely low in real life.
Practically relevant metrics

• For a commercially deployed system:
  - The True Positive rate is valid: “Out of 100 baby cries, X were detected”.
  - But False Positive rates have to be expressed per time unit: “No more than X False Alarms per year”.

• Errors translate into user experience.
  - => Need to evaluate end-to-end user experience, not just Machine Learning error rates!
Other metrics: Perf. vs computational cost

Why is computation cost important?

- The system runs “on the edge”:
  - Embedded devices
  - 10s of MIPS available
  - 100s of kilobytes of memory available

- Why not PC or cloud?
  - Cost, “bill of materials”
  - Form factor
  - Privacy!
  - Reliability
Qualitative assessment

• DET curves and EERs/DCFs give you a rate...
• ... but don’t tell you what the errors are.

Thought experiment:
• Assume a test database recorded across 100 homes with a FA rate of 20% on detecting baby cry sounds.
• Muffy the Whining Dog happens to be generating 90% of all false alarms, from a single home.
• The remaining 99% of homes share 2% of the FAs: if you ignored or solved Muffy’s single home (18% of FAs), then the FA rate would fall to 2%.
• Is this a bad system or a good system?
Qualitative assessment

• DET curves and EERs give you a rate...
• But they don’t tell you what specifically needs to be addressed to make the system better.
  ▪ In the preceding toy example, addressing dog vs baby cry confusion would solve most of the errors.
• Beware of horses!
• Are all errors “equal”?
  ▪ “Bah, it’s just the dog. It cries like a baby, doesn’t it?”
  ▪ But what if the vacuum cleaner was triggering baby cry false alarms?
Pathways to Impact

Data → Research → Lab Evaluation → Field Evaluation

Informing research topics and methodology → Impact (Application)
Field evaluation

• POLO: “Prototype Of Listening Object”, end-to-end field testing platform.
• Python UI, off the shelf hardware (Raspberry Pi)
• POLO functionality:
  ▪ Uses AA’s ai3™ to detect sounds.
  ▪ Alerts user by email.
  ▪ Supplies an audio clip of what has been detected.
• Available to research partners under contractual agreements.

Baby Crying!
Experimental design

• The POLO enables real field testing of AI³™.
  ▪ Actual TP and FA per time in the field, across a sample of representative homes.
  ▪ Method of quotas, similar to polling.

• As well as measurements of User Experience
  ▪ E.g., opinion scores.
Pathways to Impact

- Indoors data
- Taxonomy of Sounds
- Labelling cost reduction

- Machine Learning & Acoustic Features, obviously... but also
- Robustness
- Meta-parameters

- Decision
- Infinite FA priors, FA rate per time unit
- Qualitative error analysis
- Performance vs Computational Cost

- 24/7 evaluation
- User Experience

Impact (Application)
Many thanks!

By the way, we are hiring:  
http://www.audioanalytic.com

Company Information

Audio Analytic Ltd.
50 Saint Andrew’s Street,
Cambridge, UK, CB2 3AS
Web. audioanalytic.com

Tel:  +44 1223 909 305
Fax: +44 1223 750 329