**2 First Sound Event Detection System – Core Model & Thresholds**

The goal was to port the best Phase-3 classifier to the challenge setting, aggregate frame-level scores to 1.2s segments, and determine cost-optimal per-class thresholds on the development split.

**2.1 Phase-3 model recap**

We reused the 1-D convolutional network that obtained the best results during Phase 3:

• Input: 942-dimensional feature vectors per 120ms frame.

• Backbone: two Conv1d → BatchNorm → ReLU → Dropout blocks (128 filters, kernel = 5).

• Output: 58 sigmoid units.

• Training: binary cross-entropy with Adam.

**2.2 Frame-level inference**

For each development or test file we load the feature tensor T × 942, run the frozen network on GPU, and retain the 10 logits that correspond to the customer labels: Speech, Shout, Chainsaw, Jackhammer, Lawn Mower, Power Drill, Dog Bark, Rooster Crow, Horn Honk, Siren.

**2.3 Aggregation to 1.2s segments**

Following the task definition, we convert the 120ms frame scores to non-overlapping 1.2s segments (factor = 10). We tried mean, median and max pooling on the dev set; max pooling attained the lowest cost, so we chose it as the default.

**2.4 Per-class threshold sweep**

Raw segment scores are mapped to binary outputs by class-specific thresholds θ\_c. We initialize all θ\_c = 0.50 and greedily sweep θ\_c ∈ {0.05 … 0.95} while holding the other classes fixed, repeating three iterations. The objective is the customer cost, normalized to cost per minute.

**2.5 Validation cost**

After thresholding, the system reaches a total cost of 31.93 on the 20 % stratified validation split (1 648 files, 85 926 segments). The table below summarizes the class-wise contribution.

|  |  |  |  |
| --- | --- | --- | --- |
| Class | FP·min⁻¹ | FN·min⁻¹ | Cost |
| Speech | 1.89 | 0.48 | 4.31 |
| Shout | 0.84 | 0.44 | 6.10 |
| Chainsaw | 0.09 | 0.06 | 1.17 |
| Jackhammer | 0.30 | 0.13 | 2.90 |
| Lawn Mower | 0.36 | 0.11 | 2.65 |
| Power Drill | 0.82 | 0.24 | 6.02 |
| Dog Bark | 0.26 | 0.12 | 0.84 |
| Rooster Crow | 0.03 | 0.06 | 0.32 |
| Horn Honk | 0.75 | 0.22 | 5.48 |
| Siren | 0.25 | 0.09 | 2.14 |