

Rilevazione di Suoni “Anomali”

Ing. Francesco Gargiulo
Email. francesco.grg@unina.it
Tel. 081 76 83857

Palazzina DIETI - Stanza 3.20

Problema

- Riconoscimento di diverse tipologie di suoni in un contesto reale, in particolare ci concentreremo su:
 - Gun Shot
 - Scream
 - Broken Glasses

Possibili applicazioni e scenari

- Monitoraggio di zone pubbliche come:
 - Stazioni
 - Aeroporti
 - Musei
 - Piazze
 - ...
- Integrazione alla video-sorveglianza

Quali sono le difficoltà?

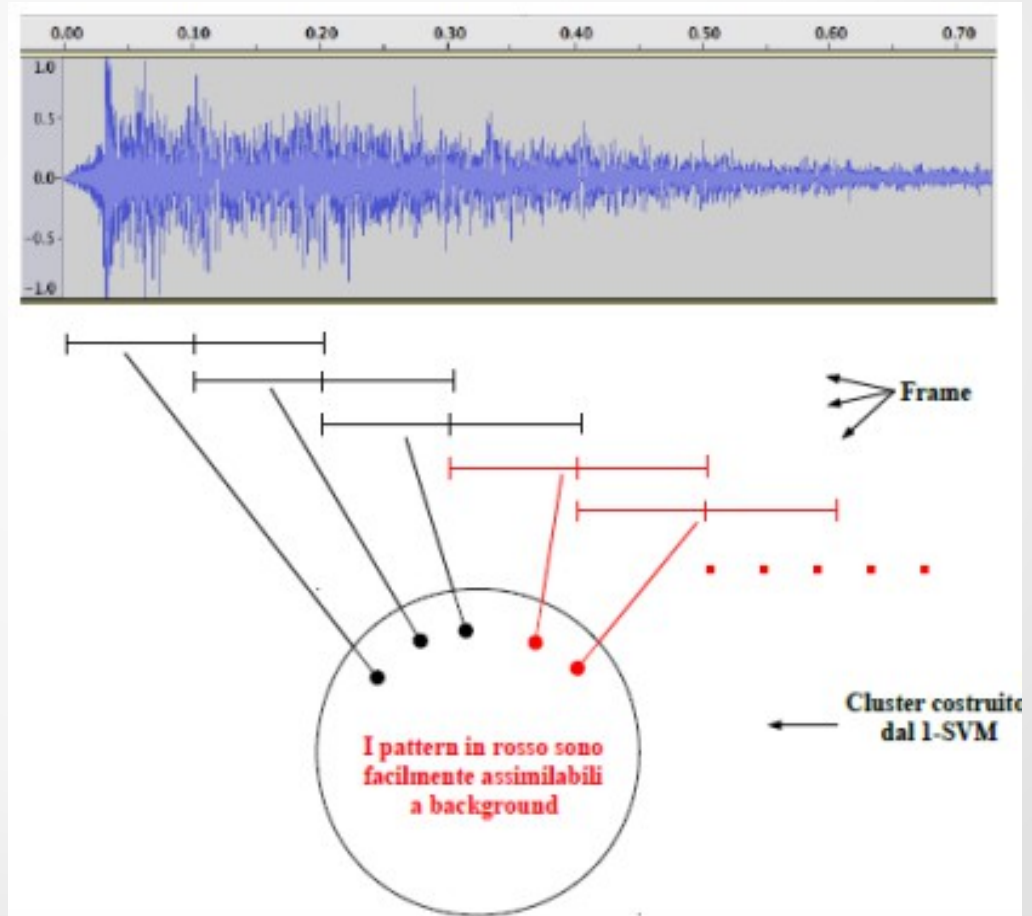
- Le diverse tipologie di suono sono molto “diverse” tra loro.
- La rilevazione dei suoni è spesso ambigua se la svincoliamo dal contesto.
- La presenza di diversi “Background Noise” complica ulteriormente il problema.

Come Affrontare il problema?

- I passi logici da seguire sono:
 - Raccogliere i dati...
 - Selezionare le feature che caratterizzano i segnali sonori.
 - Valutare se le feature selezionate sono “significative” e/o “sovrabbondanti”
 - Analizzare il problema dal punto di vista della classificazione, e definire delle strategie risolutive.
 - Definire delle metriche di performance per poter confrontare gli approcci ipotizzati.

Raccolta dei dati

- E' fondamentale avere un dataset “pulito”, che rappresenti bene il problema in esame.
 - Il caso dei GunShot:



Feature

- Alcuni possibili gruppi di feature utilizzabili sono:
 - Autocorrelation (AC)
 - Discrete Wavelet Transform (DWT)
 - Fast Fourier Transform (FFT)
 - Mel-Frequency Cepstral Coefficient (MFCC)
 - Spectral Distribution (SD)
 - Zero Crossing Rate (ZCR)

Selezione delle feature

- Questo problema può essere affrontato in due modi diversi in funzione delle scelte architetturelle che adotteremo.
 - Selezione “Globale” effettuata su tutte le tipologie di segnale.
 - Selezione “Locale” realizzata su ciascuna tipologia di segnale singolarmente.

Ipotesi-1

- Uso di un classificatore Multi-Classe
 - In questo caso la scelta di un singolo classificatore “multi-classe” è la scelta “quasi” obbligata.
 - Il vettore degli attributi da selezionare per ogni evento sonoro deve essere mappato nello stesso spazio delle feature (Selezione Globale)

Ipotesi-2

- Uso di più classificatori specializzati per i diversi suoni
 - Scelta di diversi classificatori che operano sulle diverse tipologie di suono in esame
 - Possibilità di ottimizzare i classificatori di base ottimizzandoli:
 - A livello di feature utilizzate (Selezione Locale)
 - A livello di tipologie di classificatore
 - A livello di parametri di configurazione

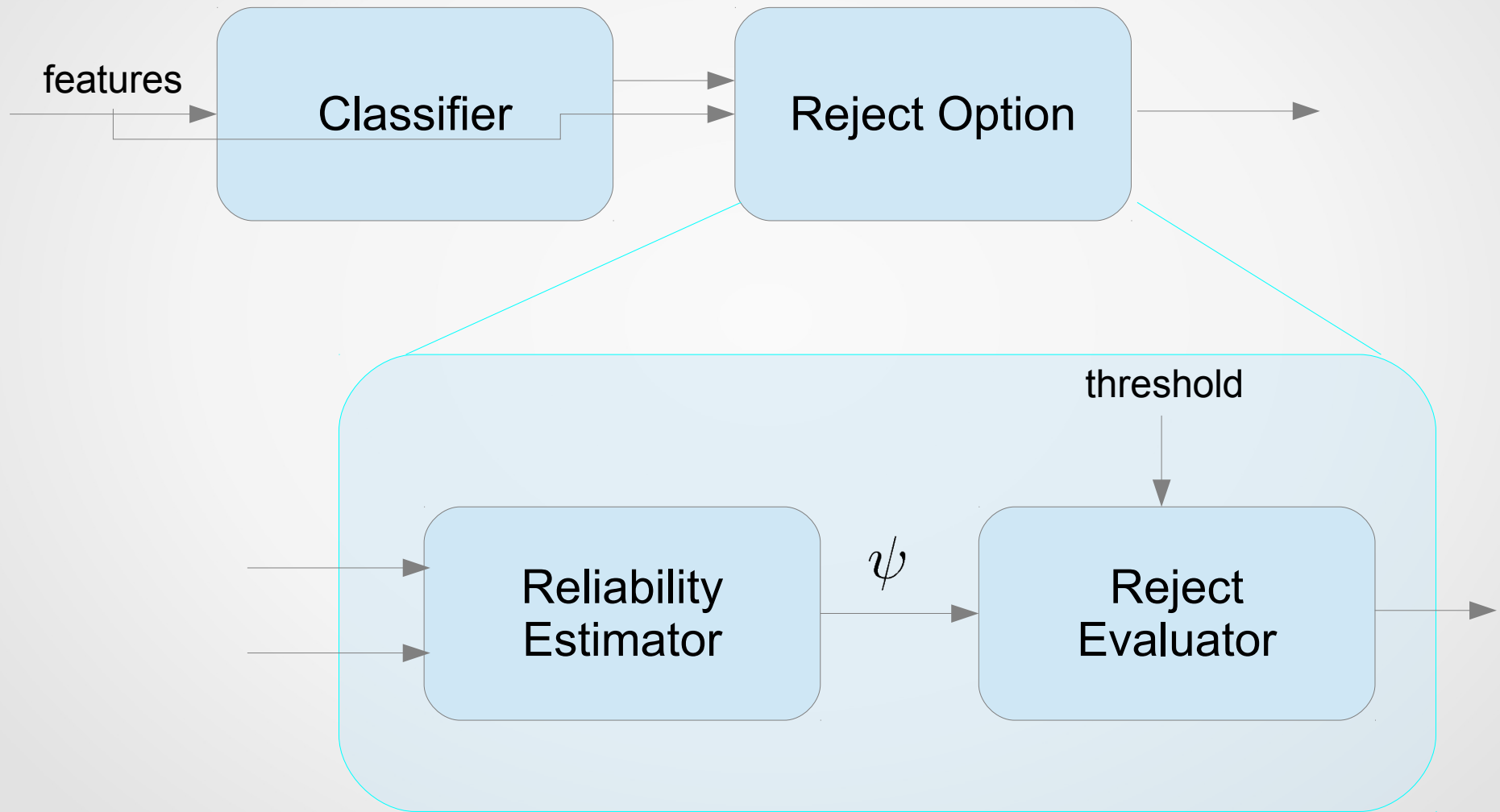
Comparazioni

APPROCCIO	PRO	CONTRO
Multi-Class.	<ul style="list-style-type: none">• Facile Implementazione	<ul style="list-style-type: none">• Bassa Modularità
One-Class.	<ul style="list-style-type: none">• Modulare	<ul style="list-style-type: none">• Prestazioni inferiori• Tuning dei parametri

Rigetto o non Rigetto???

- L'uso dell'opzione di rigetto permette di non classificare i campioni considerati “incerti”. Ma cosa presuppone?
 1. La definizione di un parametro di affidabilità associata alla classificazione.
 2. La valutazione di una o più soglie per poter definire i criteri di “rigettabilità” di un campione.

L'opzione di rigetto.



Pro e Contro del Rigetto

APPROCCIO	PRO	CONTRO
Con Rigetto	<ul style="list-style-type: none">• Maggiore affidabilità del sistema	<ul style="list-style-type: none">• Configurazione dei parametri• Gestione dei rigetti
Senza Rigetto	<ul style="list-style-type: none">• Esito su tutti i campioni	<ul style="list-style-type: none">• Minore affidabilità del sistema

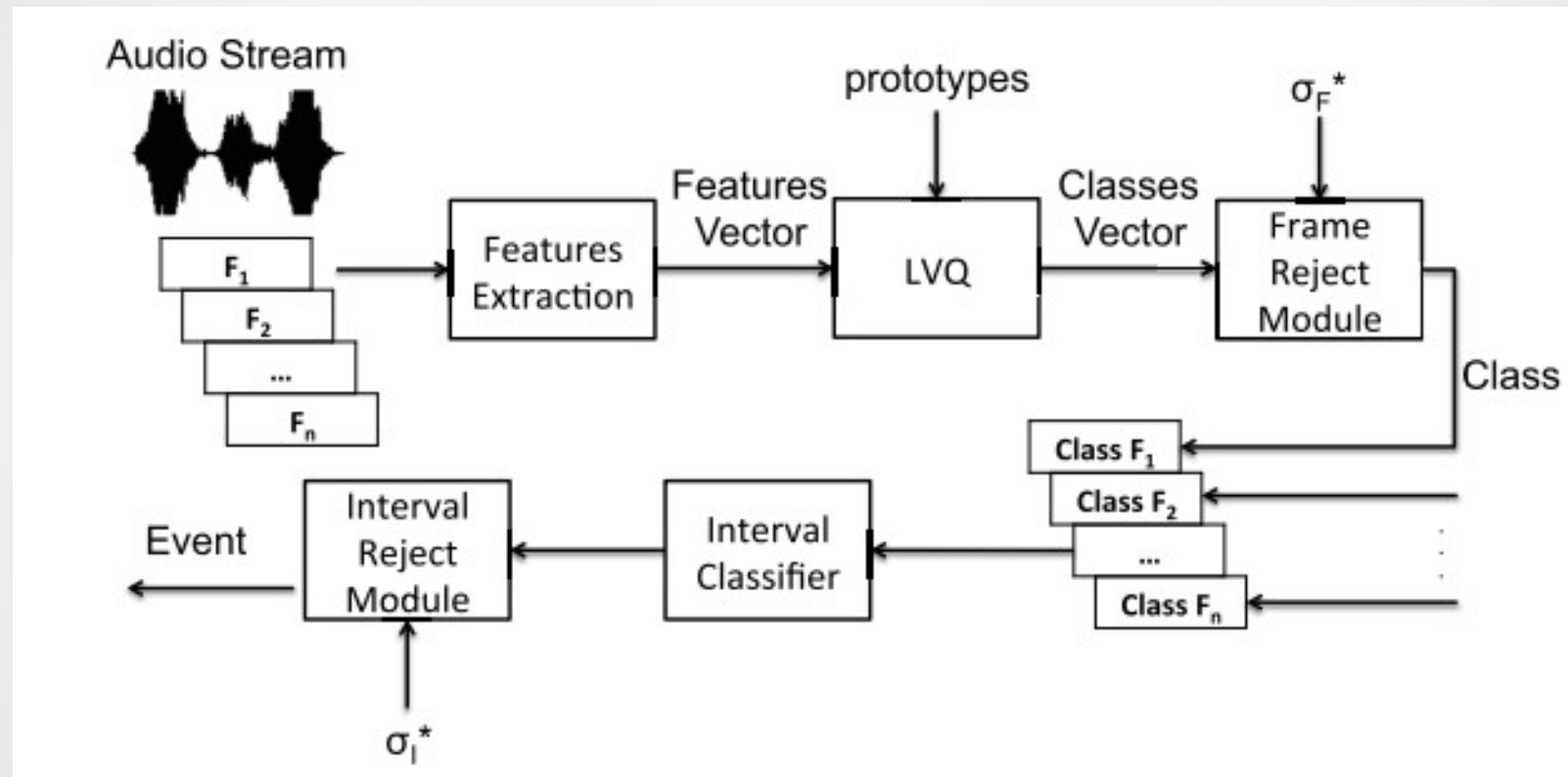
I Costi di Classificazione

- Il costo di classificazione è un concetto introdotto per diversificare il peso specifico delle diverse tipologie di errori.
- Il costo di classificazione può essere definito attraverso una matrice dei costi, o, con degli indicatori sintetici.
 - **Error Cost** (Costo associato ad un errore)
 - **Reject Cost** (Costo associato ad un rigetto)
 - **Correct Gain** (Guadagno associato alla corretta classificazione)



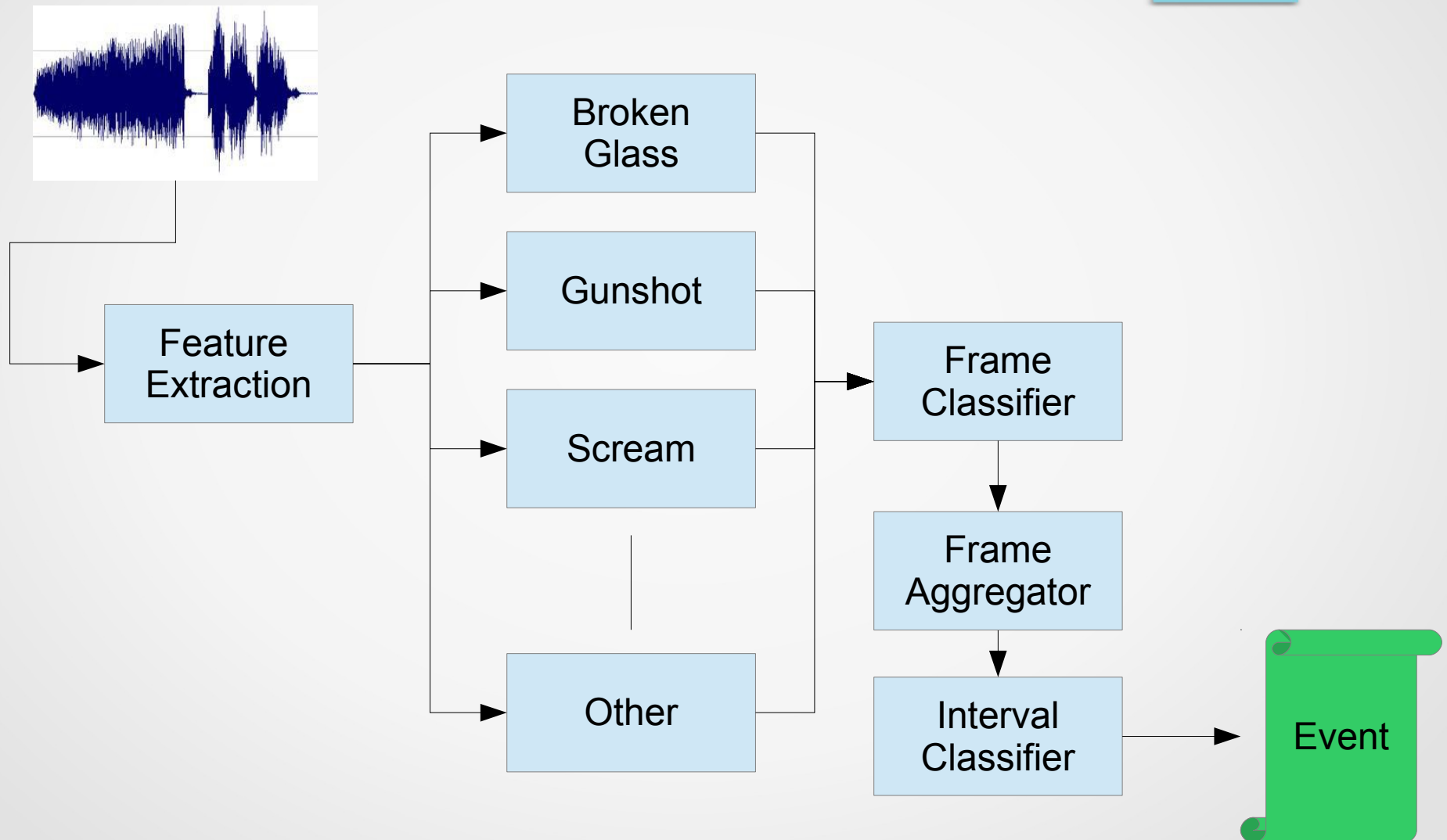
Qualche esempio “reale”

Architettura 1 “Multi-class”



* “An ensemble of rejecting classifiers for anomaly detection of audio events”, D. Conte, P. Foggia, G. Percannella, A. Saggese, M. Vento, 9th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), 2012.

Architettura 2 “One-class”



Segnali... Intervalli... Frame

- Il segnale viene suddiviso in intervalli di **1s**
- I singoli intervalli sono solitamente partizionati in frame di **200ms** sovrapposte per il **50%**.
- La frequenza di campionamento del segnale sonoro è **44kHz**
- **NOTA BENE:** Le dimensioni ottimali degli intervalli e delle frame possono dipendere dai suoni che si vogliono rilevare:
 - Gunshot (molto brevi)
 - Scream (medio-lunghi)

Dataset in esame...

	Frame (200ms;50% overlapped)	Interval (1s)
Gunshot	36	16
Scream	214	30
Glass	153	24
Background	459	59
Overall	862	129



Implementazioni in KNIME...



Domande???