RdS
RICERCA DI SISTEMA

ENEA
AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE, L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE

MINISTERO DELL'AMBIENTE E DELLA SICUREZZA ENERGETICA

Ricerca di Sistema elettrico

Tecnologie per l'analisi del contesto sonoro della smart road (LA2.23)

Prof. Vincenzo Bonaiuto

TOR VERGATA
UNIVERSITÀ DEGLI STUDI DI ROMA

RdS_PTR 22-24_PR1.7_LA2.23_375

Tecnologie per l'analisi del contesto sonoro della smart road (LA2.23)

Prof. Vincenzo Bonaiuto / Dipartimento Ingegneria Industriale – Università di Roma Tor Vergata

Dicembre 2024

Report Ricerca di Sistema Elettrico

Accordo di Programma Ministero dell'Ambiente e della Sicurezza Energetica - ENEA Piano Triennale di Realizzazione 2022-2024

Obiettivo: Decarbonizzazione

Progetto: Tema di ricerca 1.7 – Tecnologie per la penetrazione efficiente del vettore elettrico negli usi finali

Linea di attività: LA2.23

Responsabile del Progetto: Claudia Meloni, ENEA

Responsabile del Work Package: Claudia Meloni, ENEA

Responsabile Linea di Attività: Prof. Vincenzo Bonaiuto

Mese inizio previsto: luglio 2023

Mese inizio effettivo: luglio 2023

Mese fine previsto: dicembre 2024

Mese fine effettivo: dicembre 2024

Il presente documento descrive le attività di ricerca svolte all'interno dell'Accordo di collaborazione:
Tecnologie per l'analisi del contesto sonoro della smart road

Indice

1 Risultati attesi 4
2 Risultati ottenuti 5
3 Prodotti attesi 6
4 Prodotti sviluppati 7
5 Analisi degli scostamenti su attività e risultati 8
6 Sintesi delle attività svolte 9
7 Dettaglio delle attività svolte 10
7.1 Caratterizzazione delle sorgenti sonore presenti nel contesto sonoro 10
7.1.1 Sorgenti sonore di allarmi bitonali 11
7.1.2 Rumore ambientale 12
7.2 Algoritmo di riconoscimento tipologia delle sorgenti sonore 12
7.3 Algoritmo valutazione del livello di rumore ambientale 13
7.4 Interfaccia utente 14
8 Contributo delle eventuali consulenze alle attività sopra descritte 16
9 Pubblicazioni scientifiche 17
10 Eventi di disseminazione 18

Indice delle figure

Figura 1 – Microfoni installati sul veicolo elettrico messo a disposizione da ENEA 10

Figura 2 – Temporizzazione frequenze segnale bitonale per i veicoli in servizio antincendio e di urgenza ed emergenza medica 11

Figura 3 – Flow chart dell'algoritmo per l'identificazione e classificazione dei segnali SEV 12

Figura 4 – Flow chart dell'algoritmo per la classificazione e misura del livello del rumore ambientale 14

Figura 5 – Interfaccia utente.....15

1 Risultati attesi

Oggetto dell'attività di ricerca è il progetto e la realizzazione di un prototipo sperimentale per l'analisi del contesto sonoro di uno scenario urbano. Il segnale audio verrà acquisito mediante microfoni omnidirezionali installati su un veicolo elettrico messo a disposizione da ENEA.

Il risultato di tale analisi ha un duplice scopo:

1. Realizzare una mappatura dell'inquinamento acustico dell'area circostante il veicolo
2. Identificare specifiche sorgenti sonore associate a potenziali situazioni di pericolo che si trovano al di fuori del campo visivo del conducente. Ad esempio, il sopraggiungere di veicoli di emergenza con segnalazione acustica bitonale attivata.

L'informazione ottenuta dal sistema dovrà essere fornita contemporaneamente ed in tempo reale al guidatore e trasmessa alla smart road per attivare ulteriori comportamenti specifici.

2 Risultati ottenuti

L'attività svolta ha riguardato lo sviluppo di un sistema software prototipo basato su algoritmi di machine learning per l'analisi automatica del contesto sonoro di uno scenario urbano. In particolare, è stato sviluppato un sistema software sperimentale che permette di identificare e caratterizzare in tempo reale specifiche sorgenti sonore di interesse durante la marcia di un veicolo elettrico messo a disposizione da ENEA.

Il sistema sviluppato è in grado, a partire da una o più tracce audio acquisite con microfoni omnidirezionali montati sulla vettura, di effettuare:

- La mappatura del livello di inquinamento acustico dell'ambiente circostante il veicolo
- La segnalazione di eventuali situazioni di pericolo quali ad esempio la presenza di un veicolo di emergenza con i segnalatori acustici bitonali attivati

L'informazione acquisita è mostrata sullo schermo del personal computer posto a bordo del veicolo utilizzando un'interfaccia grafica appositamente sviluppata e, contemporaneamente, trasmessa alla smart road per ulteriori azioni.

Il sistema è stato provato con successo all'interno del centro ENEA Casaccia sulla vettura elettrica messa a disposizione per la ricerca.

3 Prodotti attesi

Rapporto tecnico sull'attività svolta e prototipo del sistema software per l'analisi del contesto sonoro completo di manuale d'uso.

4 Prodotti sviluppati

L'attività di ricerca ha prodotto come risultato un software che sfrutta algoritmi di Machine Learning per fornire in uscita, a partire da tracce audio acquisite dai microfoni montati a bordo della vettura elettrica:

- il livello di inquinamento acustico dell'ambiente circostante il veicolo
- La segnalazione di eventuali situazioni di pericolo quali ad esempio la presenza di un veicolo di emergenza con i segnalatori acustici bitonali attivati

Tali output sono forniti al guidatore tramite un'interfaccia di tipo visuale e al sistema di smart road per attivare ulteriori azioni.

5 Analisi degli scostamenti su attività e risultati

Non vi sono stati scostamenti, sia per quanto riguarda le attività che i risultati ottenuti, rispetto a quanto indicato nel capitolato.

6 Sintesi delle attività svolte

Obiettivo delle attività di ricerca è il progetto e la realizzazione di un software prototipo per l'analisi del contesto sonoro urbano al fine di ottenere il livello di inquinamento acustico dell'ambiente circostante il veicolo e segnalare al guidatore l'eventuale presenza di situazioni di pericolo quali ad esempio l'approssimarsi di veicoli di emergenza con i segnalatori acustici bitonali attivati.

Il software sviluppato elabora le tracce audio acquisite dai microfoni montati a bordo della vettura elettrica messa a disposizione da ENEA e, sfruttando algoritmi di Machine Learning, fornisce in uscita gli output richiesti. Tali output sono visualizzati in tempo reale al guidatore tramite un'apposita interfaccia e al sistema di smart road per attivare ulteriori azioni.

Il sistema è stato provato con successo all'interno del centro ENEA Casaccia sulla vettura elettrica messa a disposizione per la ricerca.

7 Dettaglio delle attività svolte

Oggetto del contratto è la realizzazione di un prototipo sperimentale per l'analisi del contesto sonoro della smart road, completo di driver, software e documentazione tecnica.

L'individuazione precoce della situazione di pericolo è fondamentale per gestire i comportamenti di un veicolo a guida autonoma ma anche come supporto al guidatore nell'individuazione la direzione di provenienza del veicolo di emergenza con il dispositivo acustico (segnali di allarme bitonali) in funzione. È da tenere presente che la capacità di localizzare rapidamente una fonte sonora è resa più complessa in ambiente urbano a causa delle riflessioni e riverberazioni del suono dovute agli edifici circostanti e alla carrozzeria della vettura. Inoltre, è una capacità che tende a peggiorare con l'avanzare dell'età e con l'insorgere di ipoacusie. Tenendo inoltre in considerazione che anche la velocità di reazione del guidatore è una capacità che tende a peggiorare con l'età e le patologie ad essa legata, è chiaro come la precoce identificazione e localizzazione della situazione di pericolo risulti di enorme interesse in ambito sicurezza alla guida soprattutto nel caso siano coinvolti mezzi di soccorso che presentano tipologie di guida contraddistinte da velocità elevate e traiettorie che sono necessariamente non usuali e quindi tipicamente poco prevedibili.

L'oggetto delle attività di ricerca ha riguardato l'analisi del contesto sonoro di uno scenario urbano. In particolare, è stato sviluppato un sistema software prototipale per l'analisi dell'informazione sonora di uno scenario urbano mediante l'analisi del segnale sonoro acquisito attraverso microfoni omnidirezionali posizionati su un veicolo elettrico messo a disposizione da ENEA (Figura 1).

Figura 1 – Microfoni installati sul veicolo elettrico messo a disposizione da ENEA

Lo scopo della ricerca è di mappare il livello di inquinamento acustico circostante correlandolo con le informazioni di geolocalizzazione ottenibili dall'elettronica disponibile del veicolo e segnalare eventuali situazioni di pericolo come, ad esempio, presenza e direzione di avvicinamento di veicoli di emergenza (sicurezza o soccorso pubblico) con il proprio dispositivo acustico di segnalazione attivato. L'informazione ottenuta dal sistema software prototipo sviluppato sarà resa disponibile in tempo reale sia al guidatore del veicolo che inviata alla smart road per ulteriori azioni.

7.1 Caratterizzazione delle sorgenti sonore presenti nel contesto sonoro

La prima fase delle attività di ricerca ha riguardato la caratterizzazione delle sorgenti sonore presenti nel contesto urbano di interesse per lo studio.

7.1.1 Sorgenti sonore di allarmi bitonali

In primo luogo, si è focalizzata l'attenzione su segnali di allarme bitonali dei veicoli di soccorso che possono essere raggruppati in queste tre categorie:

1. Veicoli per il servizio di urgenza ed emergenza medica (autoambulanze e auto mediche)
2. Veicoli in servizio antincendio e protezione civile (Vigili del Fuoco e protezione civile)
3. Veicoli delle forze di polizia (polizia di stato, carabinieri, guardia di finanza)

Dallo studio è emerso che i segnali emessi dal dispositivo acustico supplementare di allarme di cui sono provvisti i veicoli di soccorso sono caratterizzati da un segnale acustico diverso da nazione a nazione. Sul territorio italiano, tale segnale acustico è un segnale costituito da due toni emessi, a differente frequenza, consecutivamente e con diverse temporizzazioni. In particolare, questa sequenza è differente per i veicoli adibiti a servizi di soccorso (D.M. 17 ottobre 1980) rispetto a quella in uso su veicoli adibiti a servizi di polizia (D.M. 20 Marzo 1979).

In particolare, tali segnali sono caratterizzati da sequenze di toni alla frequenza di:

a) 392 Hz (Sol naturale) e 660 Hz (Mi naturale) per i veicoli in servizio antincendio e di urgenza ed emergenza medica
b) 466 Hz (La#) e 622 Hz (Re#) per i veicoli in servizio delle forze di polizia.

Su questi valori, la normativa prevede sia consentita una tolleranza in più o in meno del 5%.

Queste frequenze sono alternate secondo una temporizzazione all'interno di un ciclo la cui durata complessiva deve essere, per entrambi i gruppi, pari a tre secondi (± 0.5 s). In entrambi i casi i suoni emessi dai dispositivi devono susseguirsi senza interruzioni apprezzabili e senza sovrapposizioni. Nel caso di mezzi di tipo a), il ciclo completo è composto da una sequenza di toni temporizzata come mostrato nella seguente Figura 2.

392 Hz	660 Hz	392 Hz	660 Hz	392 Hz	660 Hz	392 Hz	660 Hz
1 s	160 ms	160 ms	160 ms	1 s	160 ms	160 ms	160 ms
3 s							

Figura 2 – Temporizzazione frequenze segnale bitonale per i veicoli in servizio antincendio e di urgenza ed emergenza medica

Per il gruppo b), la sequenza è composta dalla ripetizione di toni a frequenza 466 Hz e 622 Hz. Per entrambi i gruppi, fra un ciclo completo ed il successivo è consentito un intervallo di tempo la cui durata massima non deve superare i 200 ms. Anche in questo caso, la normativa ammette una tolleranza di più o meno il 5% su tutte le frazioni di tempo indicate. Per quanto riguarda il livello di pressione sonora globale per ciascuna delle note emesse dai dispositivi, questo deve essere massimo sul suo asse principale e compreso fra i 115 e i 125 dB se misurato ad una distanza pari a due metri da esso.

Inoltre, tali dispositivi sono, allo scopo di ridurre l'inquinamento acustico, estremamente direzionali e, proprio a causa della loro orientazione sul tetto del veicolo di emergenza, sono percepite meglio quando questo si trova dietro o di fronte gli altri veicoli e meno in direzione laterale. Questa caratteristica rende più difficoltosa l'individuazione della potenziale situazione di pericolo in prossimità delle intersezioni stradali da parte di quei veicoli che sopraggiungono dalle vie laterali.

7.1.2 Rumore ambientale

Il livello di rumore ambientale è definibile come il livello di rumore prodotto complessivamente da tutte le sorgenti di rumore presenti in un dato luogo e in uno specifico intervallo di tempo. La misura di questo valore deve essere confrontata con il valore limite assoluto di immissione sonora specificato dalla normativa vigente (D.P.C.M. 14 novembre 1997 "Determinazione dei valori limite delle sorgenti sonore"). Per valutare l'entità dei livelli acustici presenti nell'area che si vuole analizzare si vuole predisporre un'apposita campagna di misure fonometriche per ottenere una mappatura di tale livello sonoro correlata alle informazioni di geolocalizzazione della vettura.

7.2 Algoritmo di riconoscimento tipologia delle sorgenti sonore

L'algoritmo sviluppato ha come obiettivo quello di creare un sistema in grado di riconoscere la presenza di segnali acustici emessi dai veicoli di emergenza (SEV) indipendentemente dalla distanza della fonte del suono. Il sistema ha come ingresso una o più tracce audio acquisite tramite i microfoni installati sul veicolo e ciascuna di esse viene segmentata in intervalli di due secondi ciascuna con possibilità di overlap fino al 50%. Nella fase di preelaborazione ogni traccia è convertita in spettrogrammi standardizzati che sono elaborati utilizzando un primo modello di rete neurale (Modello 1) che fornisce in uscita la probabilità che tale segnale appartenga o meno alla classe SEV. In caso affermativo, lo stesso spettrogramma viene analizzato utilizzando un secondo modello di rete neurale (Modello 2) in grado di classificare la sorgente come vicolo in servizio di polizia o in servizio di emergenza (ambulanza o vigli del fuoco).

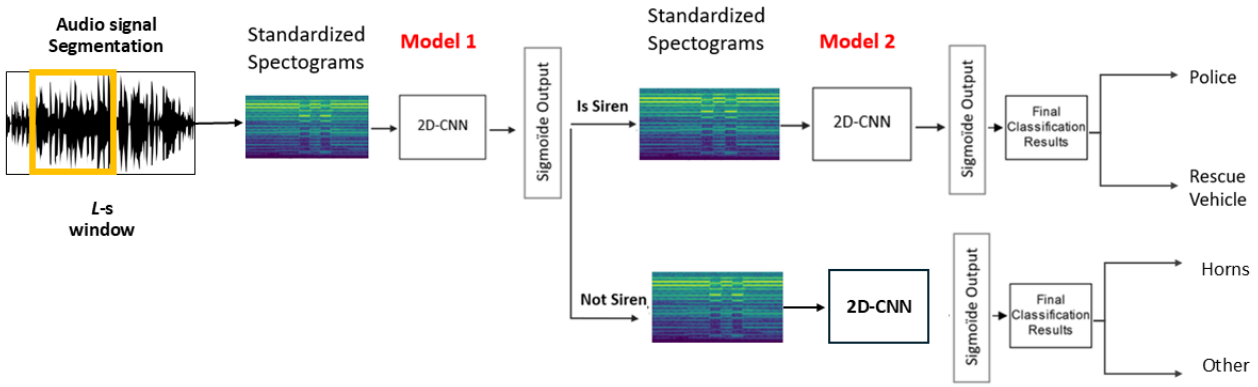


Figura 3 – Flow chart dell'algoritmo per l'identificazione e classificazione dei segnali SEV

L'algoritmo proposto fa uso di Reti Neurali Convoluzionali (CNN) che sono reti particolarmente adatte all'analisi di immagini, come sono gli spettrogrammi, grazie alla loro capacità di apprendere automaticamente e in modo adattivo gerarchie spaziali di caratteristiche attraverso i layer convoluzionali. Questi layer applicano filtri ai dati di ingresso per estrarre modelli locali che vengono poi combinati ed elaborati attraverso i layer successivi per poter riconoscere strutture complesse.

L'addestramento delle reti è stato realizzato avendo selezionato un ampio set di tracce audio acquisite in ambiente urbano in presenza di traffico automobilistico al fine di garantire che l'algoritmo possa gestire diversi possibili scenari. Nelle tracce audio proposte alla rete durante le operazioni di training sono presenti una serie di rumori comuni del traffico, tra cui clacson, presenza di animali o persone e forti rumori di motore per fare in modo che la rete possa imparare a identificare il suono dei segnali di emergenza all'interno di comuni contesti sonori urbani. Questi segnali acustici di emergenza sono stati presentati alla rete con diversi valori di

ampiezza per simulare le diverse distanze del veicolo di emergenza rispetto al ricevitore. Sono inoltre state proposte alla rete segnali ad ampiezza crescente o decrescente per simulare veicoli in avvicinamento o in allontanamento.

Terminata la procedura di addestramento della rete, sono state valutate le sue prestazioni utilizzando un set di tracce audio diverso da quello utilizzato nell'addestramento. I risultati di questi test hanno fornito i seguenti risultati:

- Validation Accuracy: 0.990
- Recall: 0.993
- Validation Loss: 0.032

La seconda rete dedicata al riconoscimento della tipologia di sirena è ancora una rete di tipo CNN che è stata addestrata utilizzando le stesse tracce audio utilizzate per la rete precedentemente descritta con la differenza che la classe dei segnali di emergenza è stata suddivisa in due classi distinte:

- segnali di emergenza di ambulanze o veicoli dei vigili del fuoco
- segnali di emergenza veicoli in servizio di polizia

L'architettura di questa seconda rete è più complessa rispetto alla prima: essa è composta da quattro diversi layer convoluzionali ciascuno dei quali è seguito da un layer di dropout e da uno di max-pooling, rispettivamente per evitare l'overfitting e per ridurre le dimensioni spaziali dei dati. Anche in questo caso, terminata la procedura di addestramento sono state valutate le sue prestazioni utilizzando un set di tracce audio diverso da quello utilizzato nell'addestramento. I risultati di questi test hanno fornito i seguenti risultati:

- Validation Accuracy: 0.996
- Recall: 0.994
- Validation Loss: 0.015

7.3 Algoritmo valutazione del livello di rumore ambientale

L'algoritmo elabora le tracce audio acquisite dai microfoni installati sul veicolo per analizzare il livello del rumore ambientale in tempo reale. Sono proposte alla rete segmenti di tracce audio di lunghezza pari a quattro secondi ciascuna con possibilità di overlap fino al 50%. Anche in questo caso si è preferito utilizzare un approccio basato su machine learning addestrando una rete con apprendimento automatico supervisionato su dati etichettati e relativi a condizioni stradali silenziose e rumorose. In questo caso il sistema è in grado di apprendere modelli sonori complessi migliorando il rilevamento in ambienti con diverse caratteristiche sonore e risulta più adattivo e accurato rispetto ai metodi basati su soglie, soprattutto in scenari di rumore misto.

In una prima fase il sistema classifica l'ambiente in due classi principali: strada tranquilla e strada rumorosa. La prima classe rappresenta le aree con bassi livelli di rumore quali le strade residenziali o aree a basso livello di traffico mentre la seconda classe corrisponde ad ambienti con livelli di rumore da moderati ad alti, tipicamente in aree urbane con traffico regolare. In una fase successiva, è stata aggiunta una terza classe per dettagliare maggiormente la classe strada rumorosa e inserendo la classe di traffico intenso (es. aree ad alta congestione di traffico).

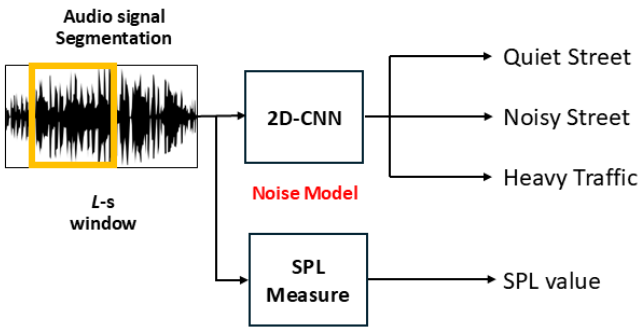


Figura 4 – Flow chart dell'algoritmo per la classificazione e misura del livello del rumore ambientale

L'algoritmo sviluppato utilizza una fase di pre-processing nella quale analizza segmenti di tracce audio per determinare le rispettive caratteristiche sonore utilizzando MFCC (Mel Frequency Cepstrum Coefficient) una tecnica molto diffusa per modellare le caratteristiche di segnali audio ed in particolare nella classificazione del parlato e dei suoni ambientali.

Questi coefficienti rappresentano lo spettro di potenza del suono in forma compatta, catturando sia le proprietà spettrali che temporali del segnale. Ciò si ottiene mediante:

1. Calcolo della Short-Time Fourier Transform (STFT) per determinare il contenuto in frequenza e di fase di porzioni di un segnale mentre questo cambia nel tempo.
2. Applicazione di un apposito banco di filtri su scala Mel per approssimare la percezione uditiva umana.
3. Calcolo del logaritmo dello spettro di potenza e l'applicazione della Discrete Cosine Transform (DCT) utile per rilevare le variazioni di informazione in un'immagine tra un'area e quella contigua.

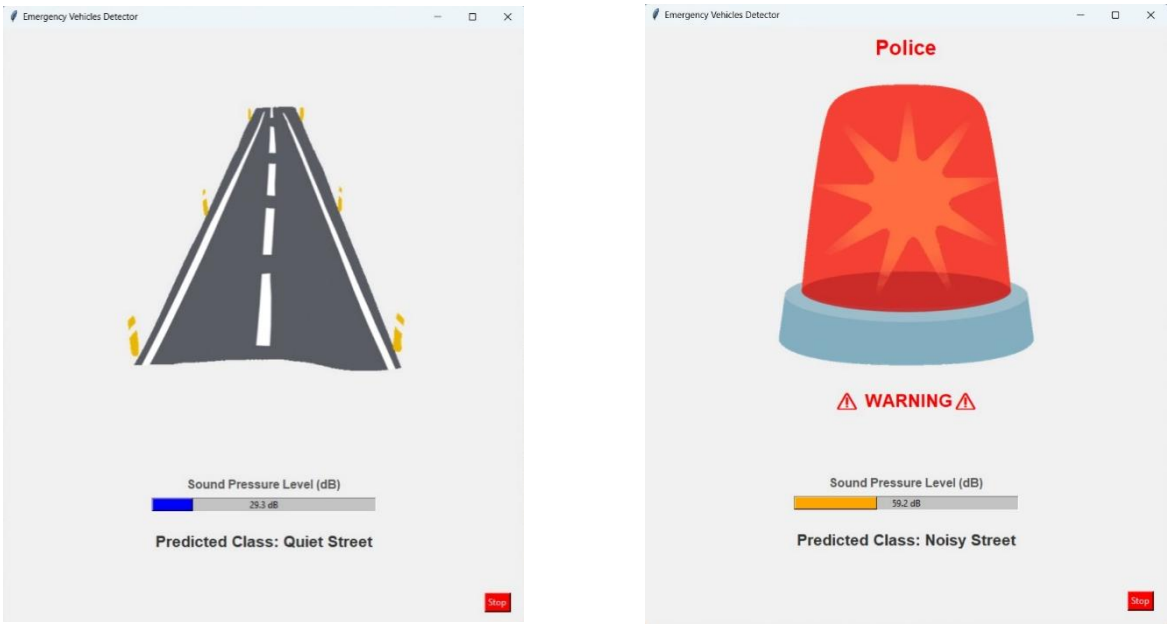
Per quello che riguarda la classificazione è stato utilizzato un classificatore basato su Support Vector Machine (SVM) che, per le sue caratteristiche, risulta essere particolarmente efficace in spazi ad alta dimensionalità e nel gestire compiti di classificazione sia di tipo lineare che non lineare. Per ogni segmento audio l'algoritmo è in grado di estrarre tredici caratteristiche MFCC insieme ai rispettivi coefficienti delta e delta-delta in grado di catturare le variazioni temporali nel suono. Il risultato finale è un vettore di caratteristiche a 39 dimensioni per ogni segmento temporale.

È stato effettuato il training del modello su un apposito dataset di tracce audio e sono stati effettuati test su un insieme di tracce non precedentemente utilizzate nella fase di training. Nella fase di valutazione delle performance il sistema ha mostrato un'accuratezza di 0,975, indicando buone capacità nella discriminazione tra le diverse classi: Strada tranquilla, Strada rumorosa e Traffico intenso.

I tempi di processamento di ogni singolo segmento audio variano fra un minimo di 32 ms fino ad un massimo di 47 ms (con un valore medio pari a 38.4 ms).

7.4 Interfaccia utente

È stata progettata un'interfaccia utente per comunicare al guidatore in tempo reale l'eventuale presenza di un veicolo di emergenza. Questa è mostrata in Figura 5 dove in a) è mostrato il risultato di una strada a basso livello di rumore ambientale (29.3 dB) e assenza di veicoli di emergenza in avvicinamento. In Figura 5b è mostrata la stessa interfaccia nel caso di presenza di un veicolo di soccorso (ambulanza o Vigili del Fuoco) in avvicinamento.



a) b)

Figura 5 – Interfaccia utente

Il sistema è stato provato con successo all'interno del centro ENEA Casaccia sulla vettura elettrica messa a disposizione.

Deliverable dell'attività di ricerca è il rapporto tecnico sull'attività svolta ed in particolare il progetto, l'implementazione ed i risultati dei test di un prototipo sperimentale per l'analisi del contesto sonoro della smart road. Sono altresì forniti i relativi driver e la documentazione tecnica comprensiva di manuale utente. Tutto il materiale è disponibile sul repository di Progetto.

8 Contributo delle eventuali consulenze alle attività sopra descritte

Non è stato necessario fare uso di consulenze esterne per realizzare le attività descritte nel presente report.

9 Pubblicazioni scientifiche

10 Eventi di disseminazione