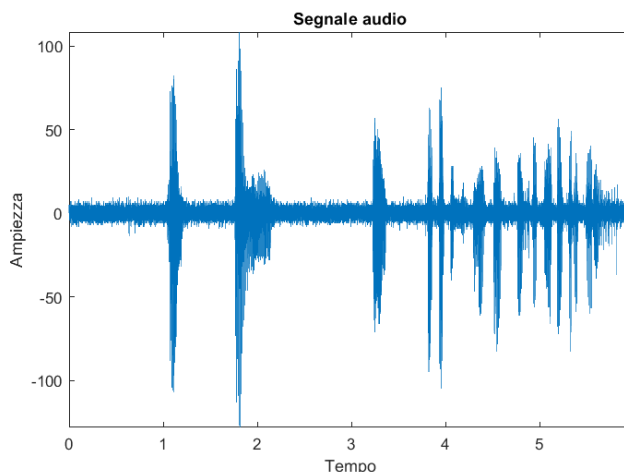


# PROGETTO DI FINE CORSO: ALGORITMO DI VAD

Ceccon Gioele 1190987

Il Voice Activity Detection (VAD), conosciuto anche come Speech Activity Detection o Speech Detection, è una tecnica usata nel processamento dei file audio che ha lo scopo di rilevare la presenza o l'assenza del parlato umano.



Per questo progetto, si ricevono in input segnali digitali audio mono, in formato PCM, da analizzare in tempo reale. Ogni segnale audio viene pacchettizzato dal trasmettitore e diviso in pacchetti da 160 byte, corrispondenti ad un intervallo audio di circa 20ms. L'algoritmo di VAD deve dunque determinare quali pacchetti hanno contenuto vocale e vanno trasmessi e quali invece possono essere soppressi in quanto non significativi. È importante notare come anche quest'ultimi pacchetti contengano rumore di fondo, diverso da 0, che viene di conseguenza sommato nei pacchetti che contengono anche voce.

## ALGORITMO PROPOSTO

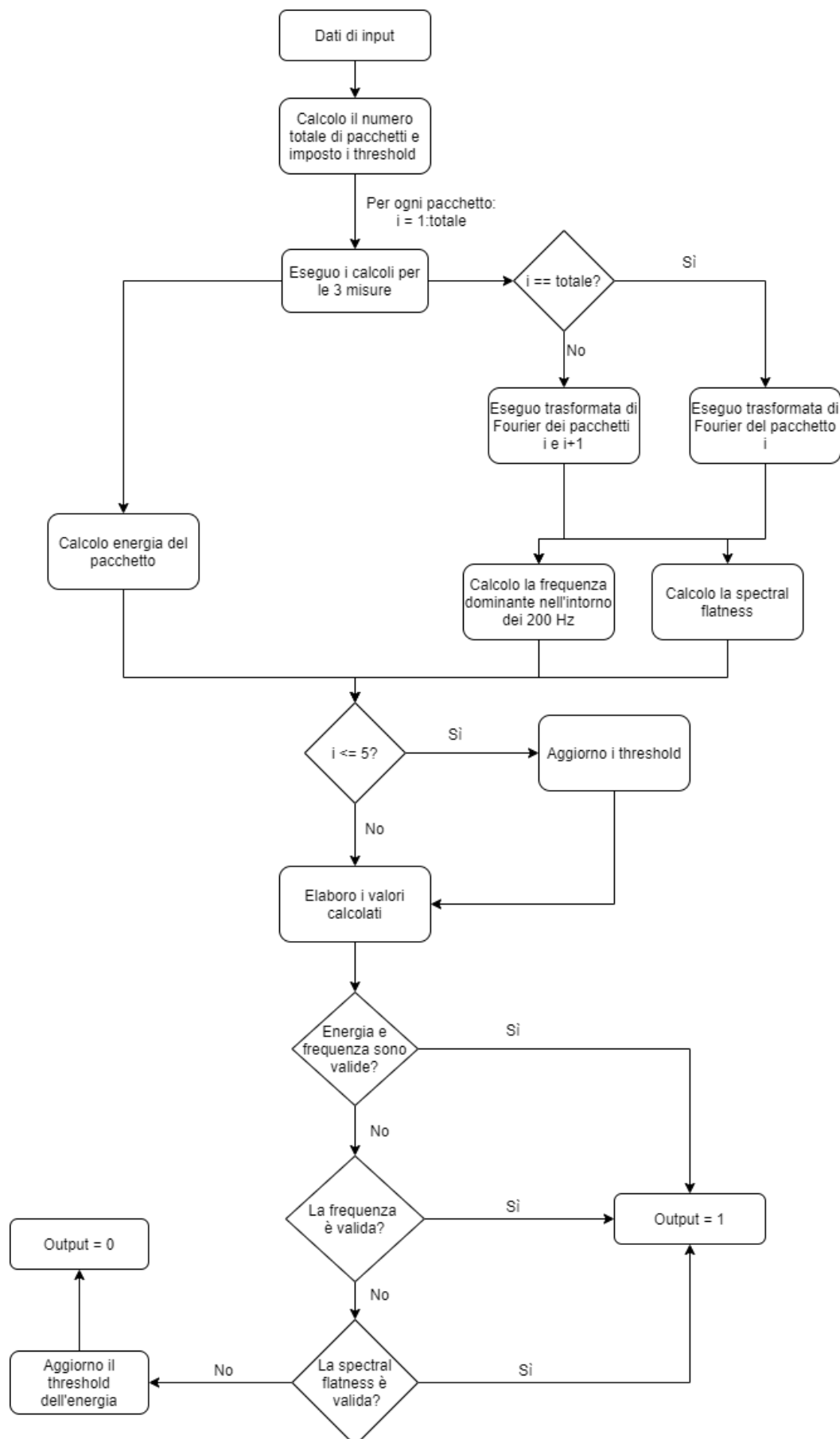
L'algoritmo proposto considera ogni pacchetto singolarmente (nel caso del calcolo dell'energia) o insieme al pacchetto direttamente successivo (per il calcolo della trasformata di Fourier) e per decidere utilizza tre misure:

- l'energia del pacchetto: regioni ad energia maggiore indicano un'alta probabilità di presenza di voce. È possibile inserire un valore limite per definire le regioni di decisione, tuttavia questa misura, utilizzata singolarmente, risulta essere non abbastanza flessibile per gestire tutte le casistiche presentate da un segnale audio
- la frequenza dominante: è la frequenza massima osservabile. In questo caso particolare, la si ottiene analizzando le frequenze vicine ai 200 Hz del pacchetto, in quanto la voce umana tende ad attestarsi su questi livelli (dunque è una frequenza dominante "relativa", circoscritta a quel range di frequenze)
- la spectral flatness del pacchetto: si tratta di una misura che caratterizza lo spettro audio. Maggiore è questo valore, maggiore è la probabilità che il pacchetto analizzato contenga voce

Queste tre misure, calcolate di volta in volta per ogni pacchetto, una volta arrivati al momento di dover prendere la decisione non sono poste sullo stesso livello. Vengono valutate infatti, in ordine, energia e frequenza dominante, frequenza dominante presa singolarmente e, per ultima, spectral flatness. Se almeno una di queste tre valutazioni risulta avere esiti positivo, l'output dell'algoritmo di VAD per il pacchetto è 1, altrimenti 0.

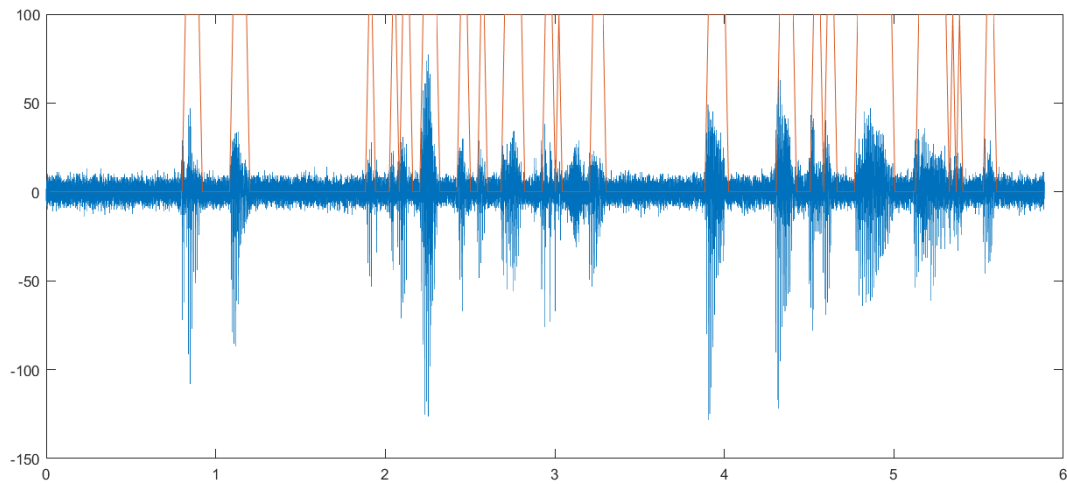
Inoltre, per ognuna di queste tre misure sono presenti dei "valori limite" (threshold in inglese) che definiscono le regioni di decisione dell'algoritmo. Questi tre threshold hanno un valore di default, impostato in seguito a prove sperimentali e ricerca sull'argomento. Per rendere l'algoritmo più adattabile alla singola traccia audio, i primi cinque pacchetti subiscono un'analisi ulteriore, volta a perfezionare i valori dei threshold.

## FLOW-CHART



## ESEMPIO DI RISULTATO

Risultato ottenuto usando come file input audio2.data



## Contenuto di outputVAD2.txt

[illegible]

## SITOGRAFIA

- Cracco Andrea, *Voice Activity Detection su segnali audio rumorosi mediante analisi Wavelet*,  
[http://tesi.cab.unipd.it/23543/1/Tesi\\_Cracco\\_Andrea.pdf](http://tesi.cab.unipd.it/23543/1/Tesi_Cracco_Andrea.pdf)
- Tom Bäckström, *Voice activity detection (VAD)*,  
<https://wiki.aalto.fi/pages/viewpage.action?pageId=151500905>
- M.Hadi, M. R. Pakravan, M. Razavi, *An Efficient Real-time Voice Activity Detection Algorithm using Teager Energy to Energy Ratio*,  
[https://www.researchgate.net/publication/334992384\\_An\\_Efficient\\_Real-time\\_Voice\\_Activity\\_Detection\\_Algorithm\\_using\\_Teager\\_Energy\\_to\\_Energy\\_Ratio](https://www.researchgate.net/publication/334992384_An_Efficient_Real-time_Voice_Activity_Detection_Algorithm_using_Teager_Energy_to_Energy_Ratio)
- M. H. Moattar, M. M. Homayounpour, *A simple but efficient real-time voice activity detection algorithm*,  
[https://www.eurasip.org/Proceedings/Eusipco/Eusipco2009/contents/papers/1569192958.pdf?source=post\\_page-----](https://www.eurasip.org/Proceedings/Eusipco/Eusipco2009/contents/papers/1569192958.pdf?source=post_page-----)