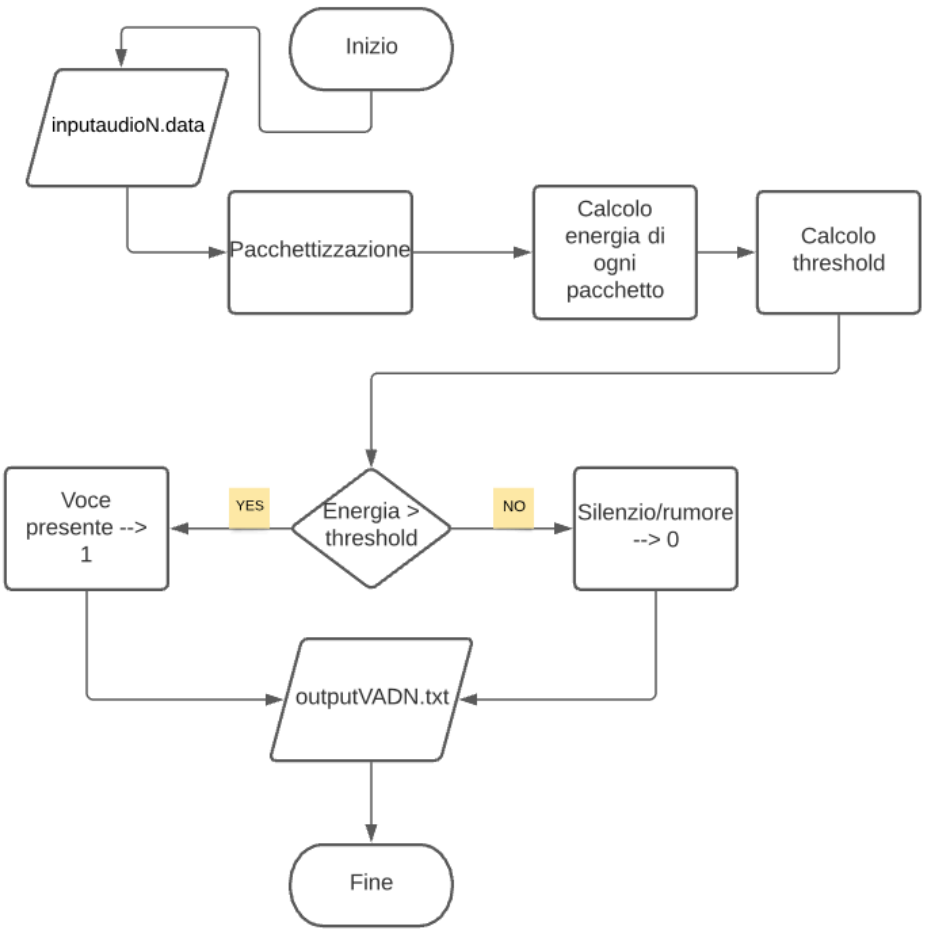
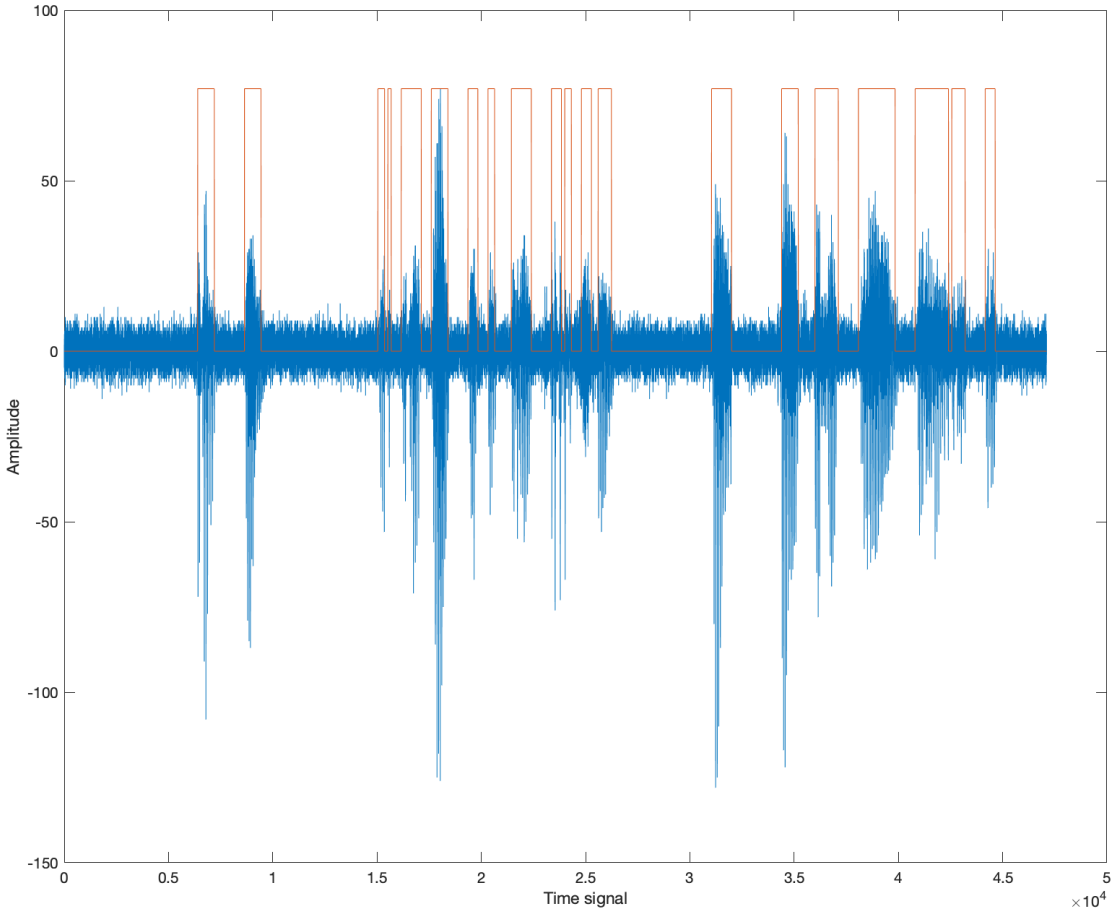
Trambaiollo Luca N. Matricola:1196021

Reti di Calcolatori
PROGETTO DI FINE CORSO
Progettazione e realizzazione algoritmo di Voice Activity Detection

Un segnale vocale non è un segnale stazionario, infatti la sua ampiezza varia nel tempo, dove il silenzio è rappresentato dallo zero. I valori del segnale vocale possono essere positivi o negativi, a seconda che la pressione d'aria provocata dall'onda sonora risulti essere superiore o inferiore alla pressione atmosferica in condizioni di silenzio. Siccome a volte parliamo energicamente e altre volte no, il segnale vocale e dunque anche l'onda sonora, ha delle variazioni. È quindi possibile utilizzare l'energia del segnale come indicatore della presenza del parlato. La voce aggiunge energia al segnale, in modo tale che le regioni ad alta energia dello stesso siano probabilmente parole. Si imposta quindi una soglia di silenzio tale che quando l'energia del segnale σ^2(x) è al di sopra della soglia, il VAD indichi l'attività del parlato. Come valore di tale soglia, dopo svariate prove sperimentali, ho scelto di prendere la media dell'energia dei primi tre pacchetti (primi 60 ms) in quanto nel 99% dei casi il primo secondo è rumore bianco standard della traccia audio (silenzio).
Quindi per ogni pacchetto, formato da 160 campioni (corrispondenti a circa 20ms), ho calcolato l'energia totale normalizzata e confrontato quest'ultima con la soglia ottenuta per vedere se assegnare 1 (voce presente) nel caso fosse maggiore, o 0 (silenzio/rumore) in caso contrario.
Ho scelto di arrotondare per difetto il numero di pacchetti nel caso in cui la divisione tra il numero totale di campioni e 160 (dimensione di ogni pacchetto) avesse resto, in quanto non ho considerato se il pacchetto non fosse completo, quindi composto da 160 campioni.
L'algoritmo utilizzato in formato flow chart è il seguente:



Esempio di output applicando l'algoritmo VAD a inputaudio2.data in cui viene mostrato come l'ampiezza del segnale varia in dipendenza del tempo (ovvero come è distribuita l'energia), nel quale appunto in blu è visibile l'ampiezza del segnale audio nel corso del tempo e in rosso la traccia che fa capire la presenza o meno di voce nei vari pacchetti (quando è uguale a 0 c'è silenzio/rumore, quando è alta è presente voce) :



L'algoritmo che ho scelto di utilizzare mi è sembrato il più adeguato per analizzare un segnale digitale audio mono, in formato PCM, che si assume essere generato in tempo reale in quanto determina al meglio quali sono i pacchetti che hanno contenuto vocale e, quindi, vanno trasmessi, e quali invece possono essere soppressi in quanto privi di contenuto vocale significativo.