

PROGETTO DI FINE CORSO**Progettazione e realizzazione algoritmo di Voice Activity Detection****1. Introduzione**

Si vuole progettare un algoritmo di riconoscimento del parlato (Voice Activity Detection - VAD) che determini quali pacchetti hanno contenuto vocale e, quindi, vanno trasmessi, e quali invece possono essere soppressi in quanto privi di contenuto vocale significativo. Per ogni pacchetto audio, l'algoritmo VAD dovrà generare in uscita il valore 1 nel caso il pacchetto debba essere trasmesso, e 0 altrimenti. L'obiettivo è trasmettere il minore numero possibile di pacchetti che consenta la ricostruzione del segnale vocale senza artefatti (clipping).

2. Note aggiuntive

- I segnali digitali audio mono forniti in input sono in formato PCM e si assumono essere generati in tempo reale.
- La voce umana ha uno spettro che si estende nell'intervallo [200; 3400] Hz.
- La decisione dell'algoritmo VAD relativamente al pacchetto n-esimo si basa anche su (fino a) 4 pacchetti precedenti e su (fino a) 2 pacchetti successivi, con un ritardo complessivo di codifica di massimo 40 ms.
- Ad ogni iterazione, per lo studio di ogni pacchetto è stato possibile visualizzare i campioni interessati nel dominio delle frequenze tramite la Trasformata di Fourier veloce (FFT) e il calcolo del relativo spettro unilaterale.
- Per essere il pacchetto n-esimo valutato come *voce*, i campioni considerati devono avere almeno un valore di ampiezza nell'intervallo [200; 3400] Hz superiore ad una determinata soglia. La soglia viene stabilita in base a prove "manuali" al fine di raggiungere il miglior compromesso di soglia per tutti gli audio di input. Soglia determinata: 0,71.
- L'algoritmo è sviluppato in linguaggio Matlab con la versione del software MATLAB R2021a.

3. Algoritmo proposto

È presente il flow chart dell'algoritmo nella pagina seguente di questo documento.

4. Esempi di risultato

Di seguito si riportano 3 esempi di file audio dati in input, prima e dopo l'applicazione del VAD. I grafici dei file audio dopo l'applicazione del VAD rappresentano i file audio di input con azzerati i pacchetti valutati come *non voce*.



