POLITECNICO DI TORINO

Corso di Laurea in Ingegneria del Cinema e dei mezzi di Comunicazione

a.a. 2017/2018

Corso di Elaborazione dell'Audio Digitale

Progetto:

Clip Alignment

PITTARO Pietro – matr. 198000

Scopo del progetto

Sostituire la traccia audio di un video con quella di un audio recorder
La traccia dell'audio recorder va sincronizzata con quella del video gestendo:

1. Differenti tempi di inizio e fine (anche qualche minuto)
2. Possibile introduzione di silenzi (es. smartphone)
3. Possibile disallineamento dei clock (deriva, ordine ms)

Attività svolta

E' stato realizzato un programma in MATLAB che dati in input i due file WAV con le tracce di riferimento calcola il file di output con la registrazione dell'audio recorder epurato di eventuali silenzi e derive sincronizzato con la traccia del video (che fa da riferimento temporale)

1. Il disallineamento iniziale viene individuato attraverso la cross correlazione. Il risultato ottenuto è corretto e la funzionalità individua sempre in modo estremamente preciso e acusticamente perfetto il disallineamento delle due tracce (positivo e negativo). Nel caso di traccia audio recorder in ritardo rispetto a quella del video il programma utilizza per questo tempo la traccia audio del video (prevista la possibilità alternativa di azzeramento). La stessa considerazione è stata fatta per la parte finale dell'audio ricostruito. L'ampiezza della finestra di ricerca del disallineamento iniziale è stata impostata a qualche decina di secondi nel caso dei miei file di test che avevano al massimo qualche secondo di disallineamento e di 5 min nel caso della lezione reale dove il programma ha individuato il disallineamento iniziale di 225.998 sec: traccia audio del video inizia 225.998 sec dopo quella del cellulare.

2. La ricerca dei silenzi è stata realizzata con una scansione delle due tracce audio riallineate adottando il seguente algoritmo di scansione:

   Si procede all'analisi delle due tracce di input partendo dall'inizio e procedendo a step che chiameremo blocchi la cui dimensione per il file audio della lezione è stato impostato a 2 min mentre per i file di prova di pochi minuti creati da me per le verifiche si sono utilizzati blocchi mediamente di 30-40 sec. Il programma esegue il calcolo della cross correlazione di una finestra di confronto al fondo del blocco che chiameremo Frame di verifica. La dimensione di questa finestra è stato mediamente di 10 sec in tutte le prove effettuale.

   Il programma esegue il calcolo della cross correlazione tra i due frame corrispondenti e:

   se inferiore ad una soglia di "perfetto allineamento" (usato nelle prove 1 e 10 msec) passa al blocco successivo => avanzo di un

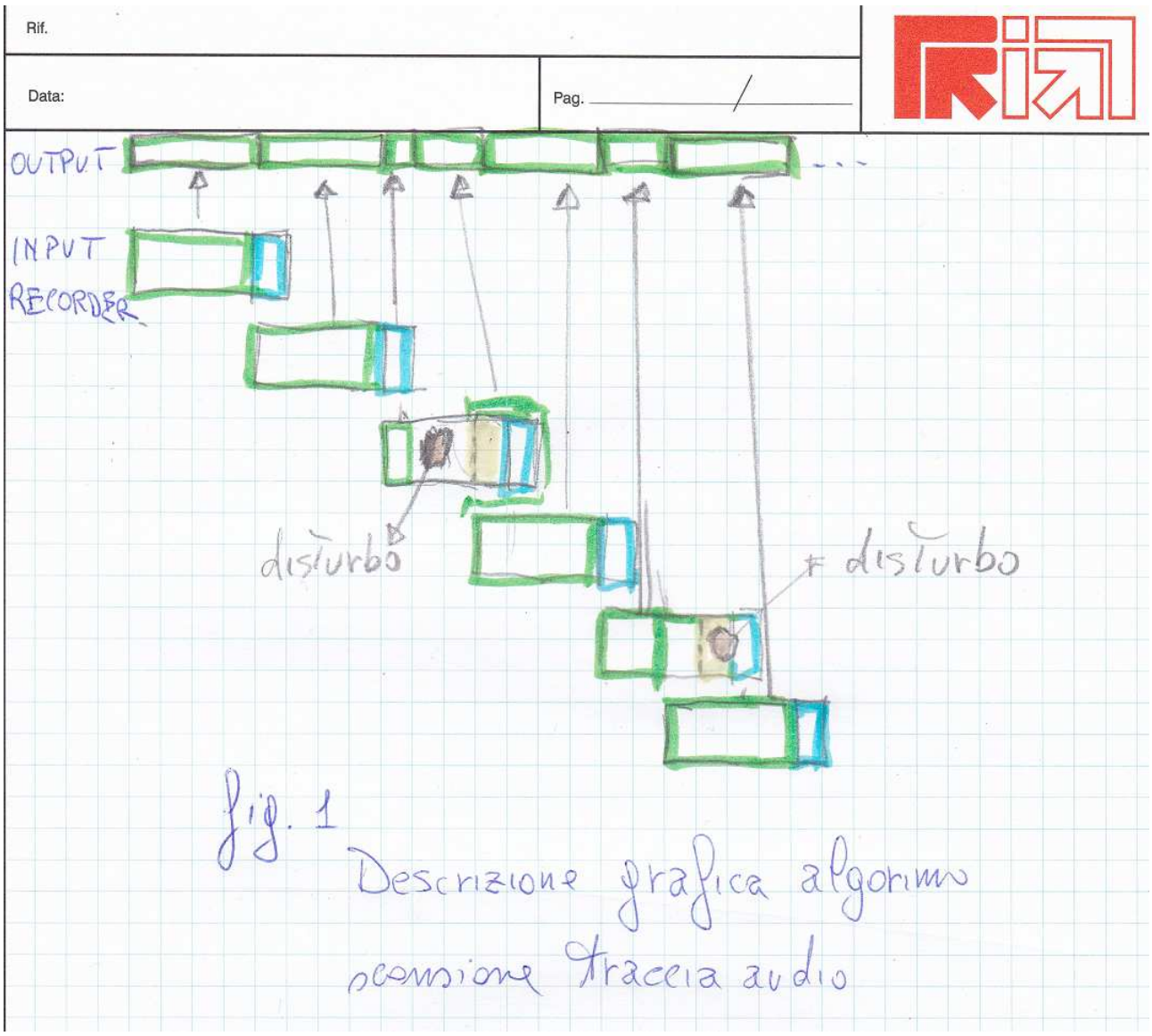
blocco a partire non dalla fine del blocco corrente ma dall'inizio del frame corrente. Misura necessaria per gestire correttamente silenzi posizionati in corrispondenza del frame

se superiore alla soglia controllo anche la cross correlazione di frame subito precedente a quello corrente:

se i due offset sono "uguali" (soglia bassa – 1 ms) allora il blocco contiene spezzone estraneo non in corrispondenza dei due frame finali del flocco e quindi eseguo la funzione di ricerca "inizio silenzio" su tutto il blocco (anche allargato all'inizio per agevolare,dipende ovviamente dal tipo di algoritmo che si utilizzerà, l'individuazione dello spezzone estraneo ad inizio blocco). La funzione se individua correttamente l'inizio del silenzio (spezzone estraneo) mi restituisce la sua posizione all'interno del blocco analizzato. Il mio programma MATLAB ha implementato in modo funzionante il solo riconoscimento di un segnale che resta sotto una soglia continuativamente per un tempo indicato. Dall'indicazione dell'istante di inizio e della durata già calcolata dalla cross correlazione il programma copia la parte di audio precedente e successiva del blocco realizzando così l'eliminazione del disturbo dalla traccia di output.

se i due offset non sono "uguali" vuol dire che lo spezzone estraneo si trova nella zona dei 2 (e anche 3) frames finali del blocco e procedo copiando la parte precedente dell'audio recorder nella traccia di output e impostando come inizio blocco successivo la "fine blocco attuale – 3 frames". Lo spezzone estraneo sarà così posizionato interamente all'inizio di questo blocco e verrà individuato al ciclo successivo (si assume che uno spezzone sia sempre unico in un blocco)

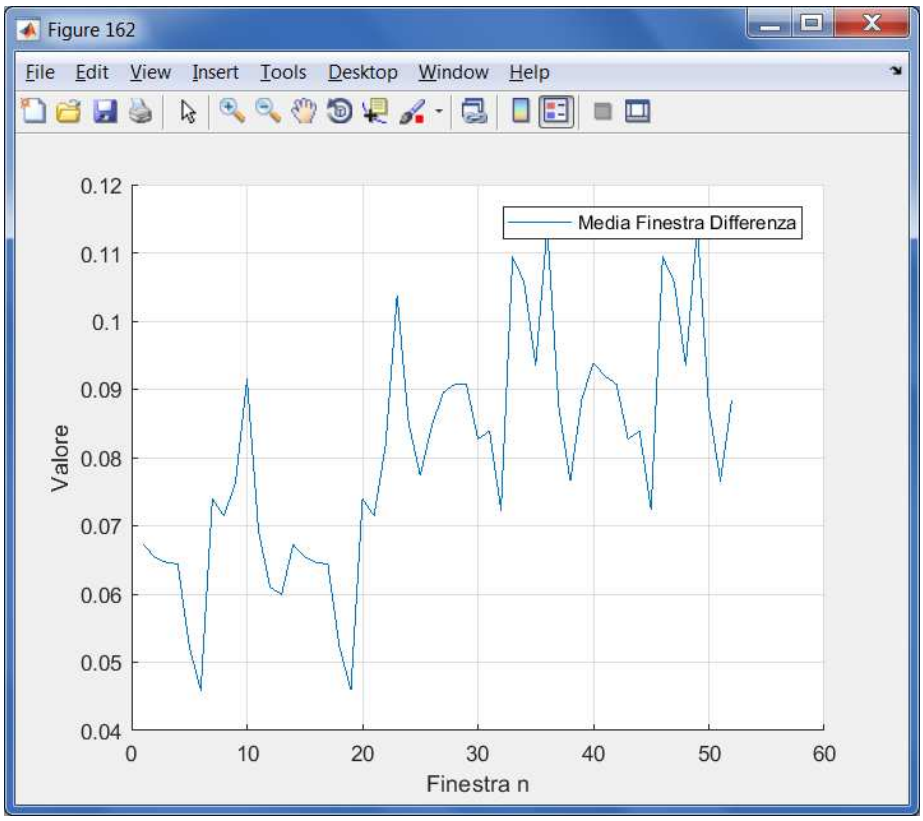
La figura che segue mostra schematicamente il tipo di algoritmo utilizzato per la scansione delle tracce audio



3. Al termine della scansione il programma verifica l'allineamento del penultimo Frame (dimensione configurabile in modo dedicato) e a seconda del segno si inseriranno o si elimineranno i campioni di deriva totale. L'operazione viene eseguita distribuendo in modo lineare l'operazione su tutta la traccia. Come nel caso di inizio file il programma gestisce l'eventuale "mancanza" della traccia audio recorder al fondo utilizzando quella della traccia video (in alternativa alla condizione di azzeramento)

NOTE:

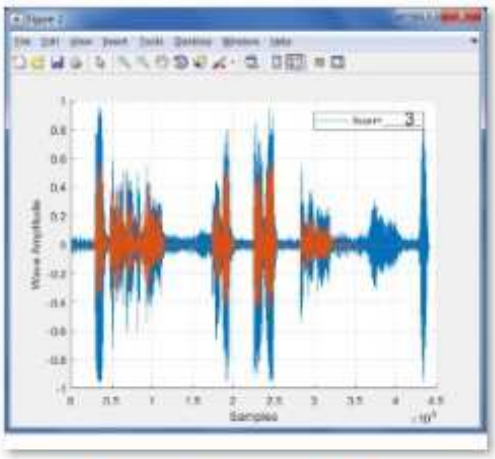
La ricerca dello spezzone estraneo (di tipo qualsiasi) è stata provata facendo la differenza dei due segnali nel blocco dove è stato individuato, ma i risultati di alcune prove non avevano dato segni evidenti indicazioni significative in corrispondenza dei 4 secondi di silenzio come mostrato dalla figura seguente che mostra l'andamento del segnale differenza in un blocco di 52 sec in cui è presente un silenzio di 4 sec



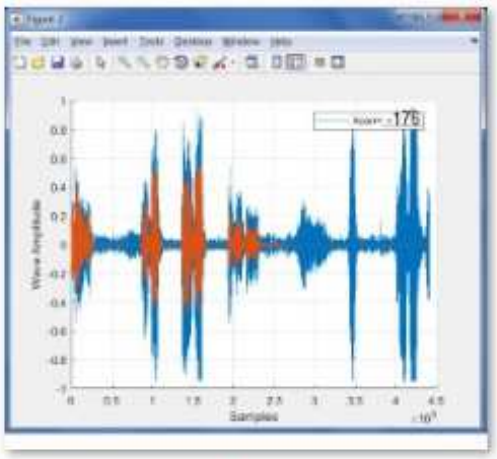
E' stata poi indagata un'altra ipotesi per l'individuazione dei segnali estranei che prevede l'utilizzo della cross correlazione e in questo caso si sono fatte alcune considerazioni interessanti che non ho però avuto modo di provare ad utilizzare in qualche implementazione reale. L'analisi dell'andamento della cross correlazione ha evidenziato che data una certa ampiezza della finestra di verifica e data una durata del segnale estraneo i valori che la cross correlazione può assumere sono sostanzialmente solo 3:

- 0
- 40 msec (176 campioni)
- 4 sec

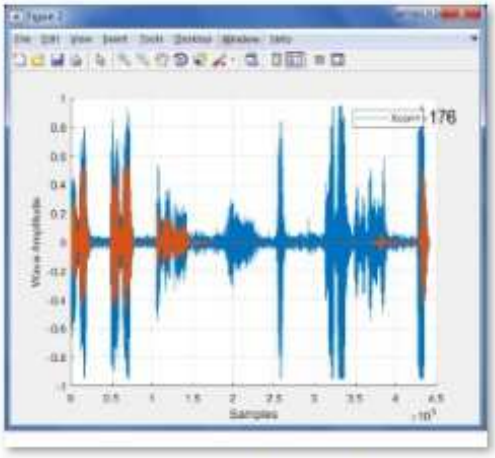
Nel mio caso di test con una finestra di verifica di 10 sec e un segnale di ampiezza 0 di 4 sec come ho potuto verificare facendo una serie di prove mirate nel posizionamento relativo delle due tracce e della finestra di verifica come mostrato nelle immagini che seguono:



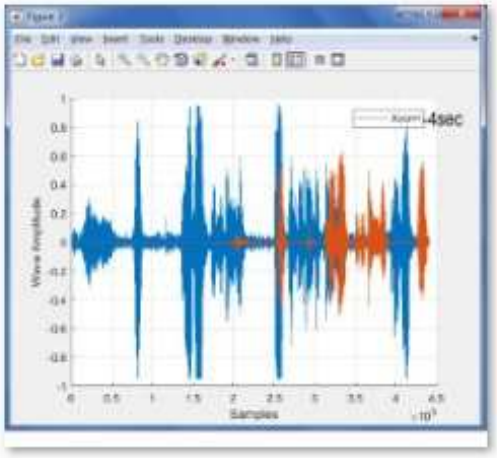
TestXcorr1-8sOK-2sSil.jpg



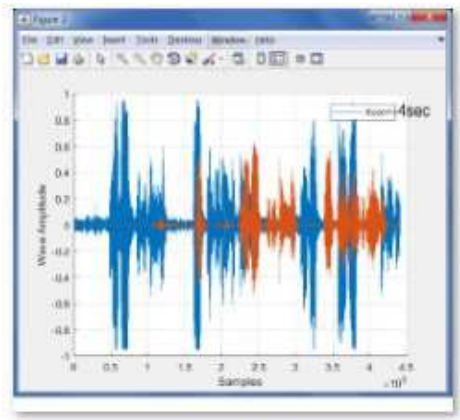
TestXcorr2-6sOK-4sSil.jpg



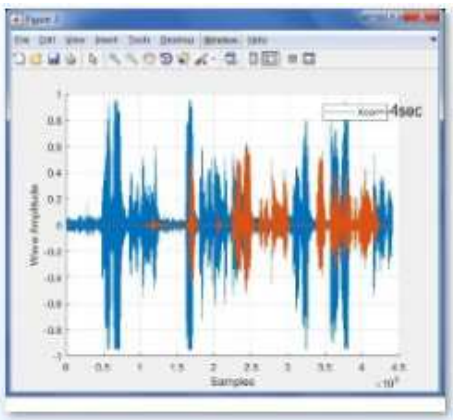
TestXcorr3-4sOK-4sSil-2sOK.jpg



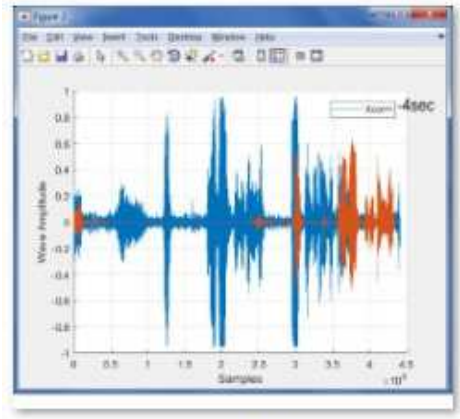
TestXcorr4-4sSil-6sS.jpg



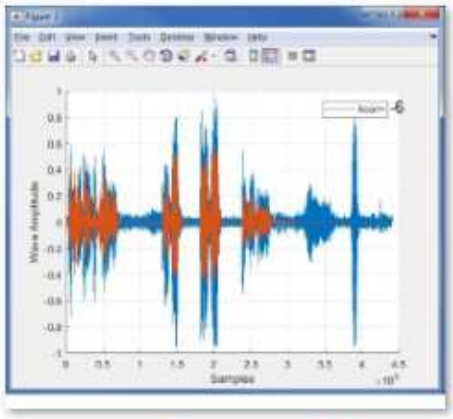
TestXcorr5-3.5sSil-6.5sS.jpg



TestXcorr6-3sSil-7sS.jpg



TestXcorr7-1sOK-4sSil-5sSh.jpg



TestXcorr8-7sOK-3sSil.jpg

Analizzando l'andamento della cross correlazione nell'intorno dei "punti di discontinuità" si è verificato che il passaggio avveniva sempre in un solo step (che è stato impostato a 10 msec) e sempre nella stessa posizione anche in punti diversi della traccia. Anche la sostituzione dell'ampiezza 0 con un segnale di un violino della stessa durata di 4 sec ha confermato che l'andamento della cross correlazione ha sempre la stessa identica forma come si può vedere dai grafici che seguono.

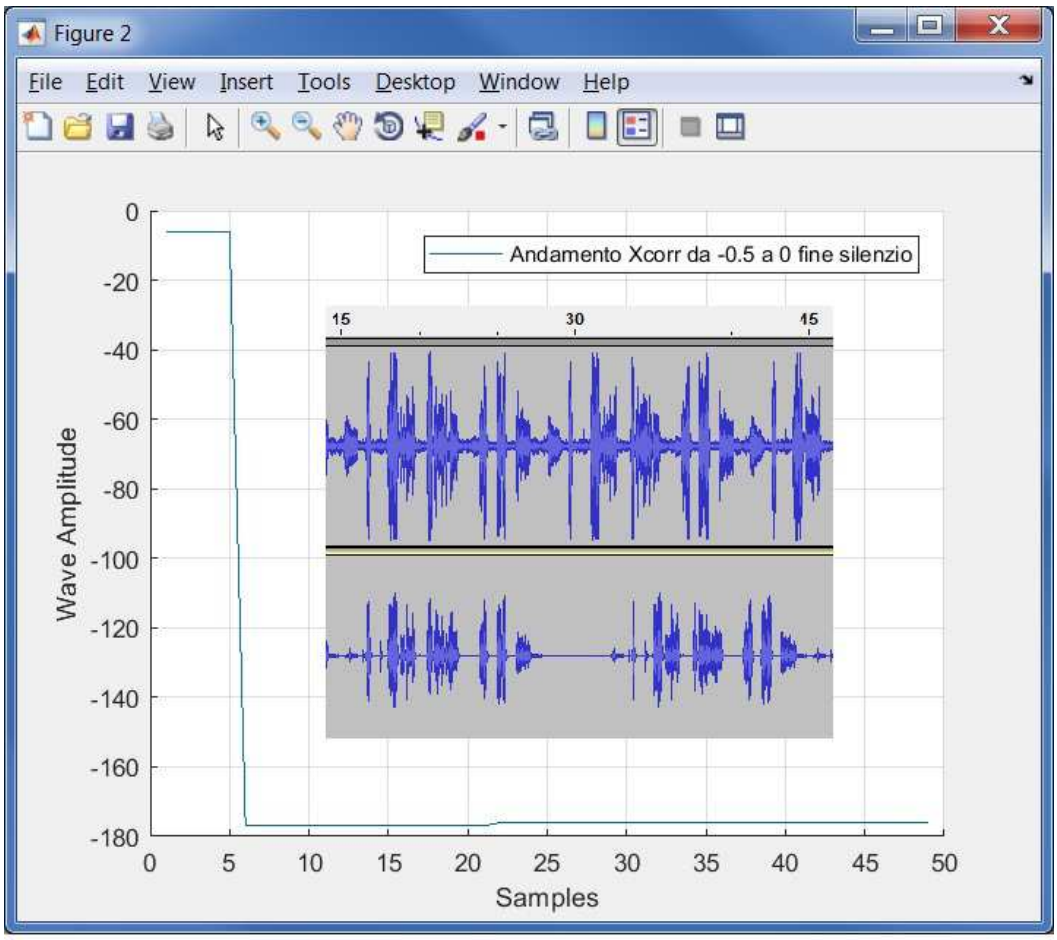


Figure 2
File Edit View Insert Tools Desktop Window Help
Andamento Xcorr da -0.5 a 0 fine silenzio
Wave Amplitude
Samples
0
-20
-40
-60
-80
-100
-120
-140
-160
-180
0 5 10 15 20 25 30 35 40 45 50
15 30 45

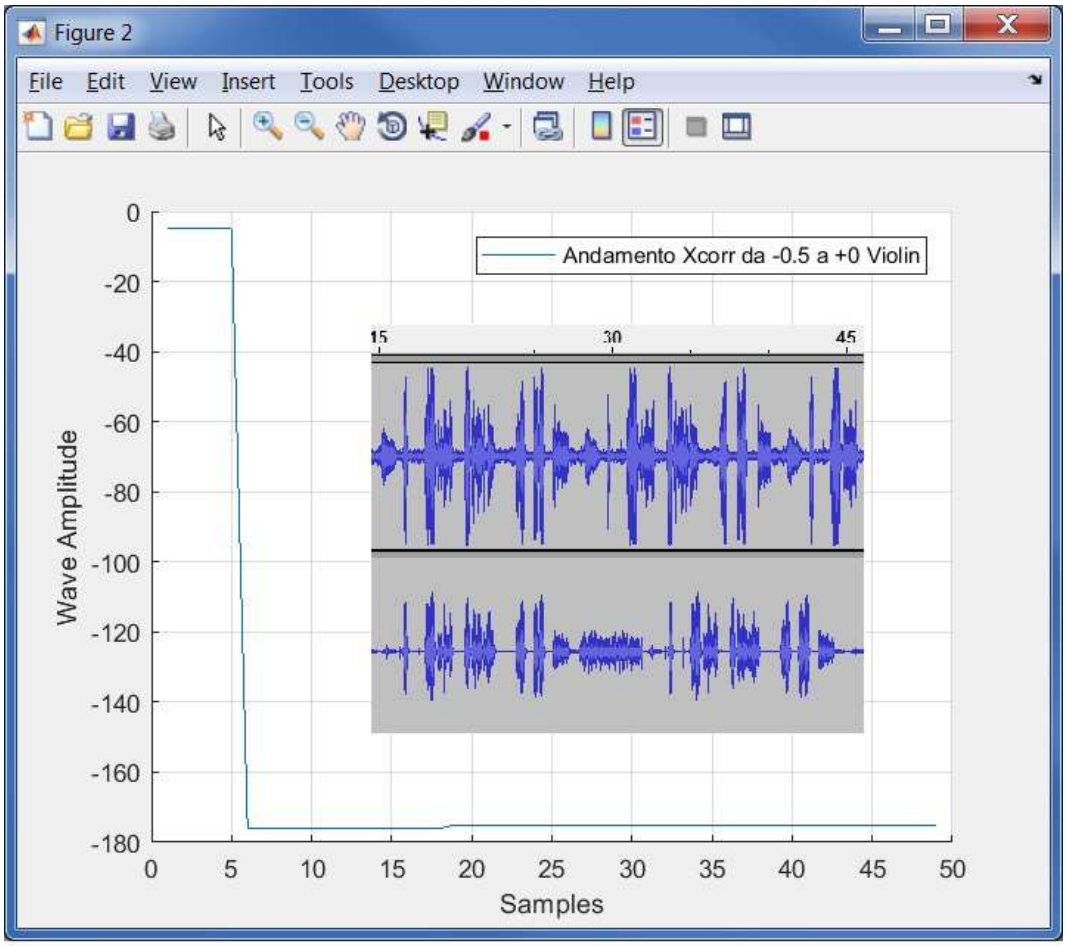


Figure 2
File Edit View Insert Tools Desktop Window Help
Andamento Xcorr da -0.5 a +0 Violin
Wave Amplitude
Samples
0
-20
-40
-60
-80
-100
-120
-140
-160
-180
0 5 10 15 20 25 30 35 40 45 50
15 30 45

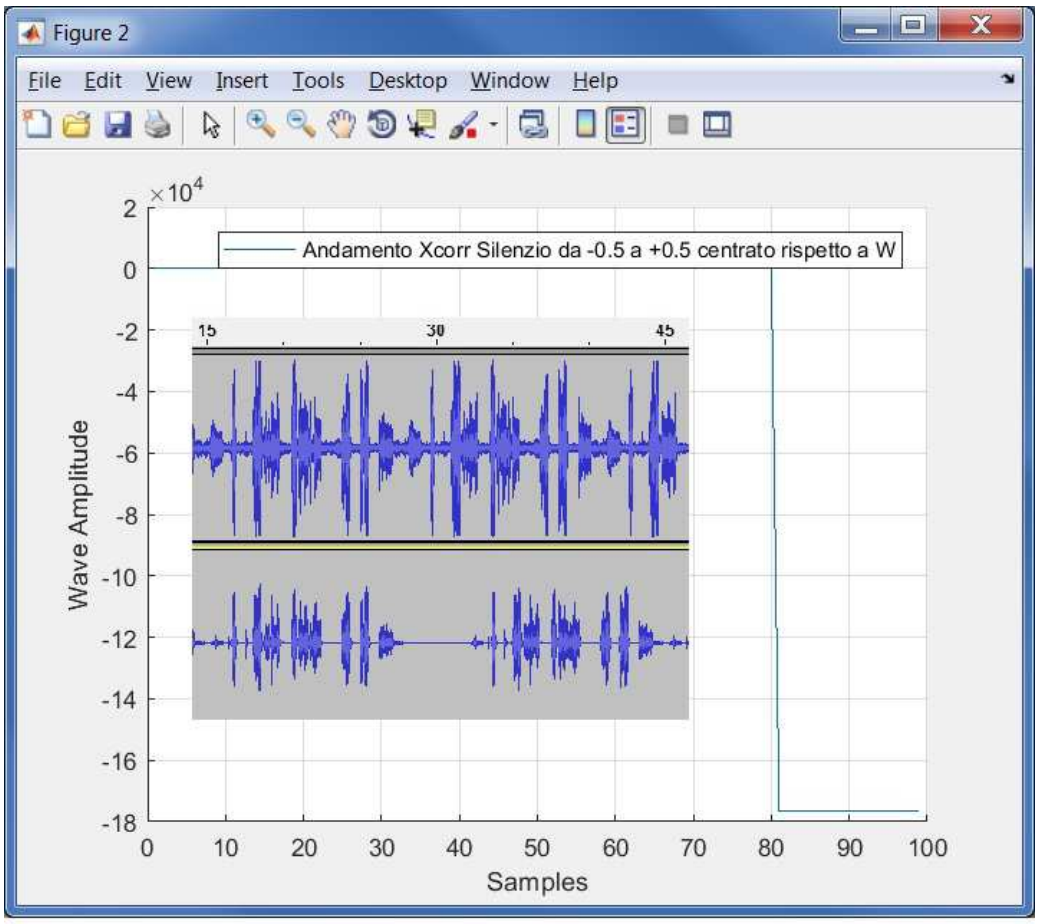


Figure 2
File Edit View Insert Tools Desktop Window Help
×10^4
Andamento Xcorr Silenzio da -0.5 a +0.5 centrato rispetto a W
Wave Amplitude
Samples

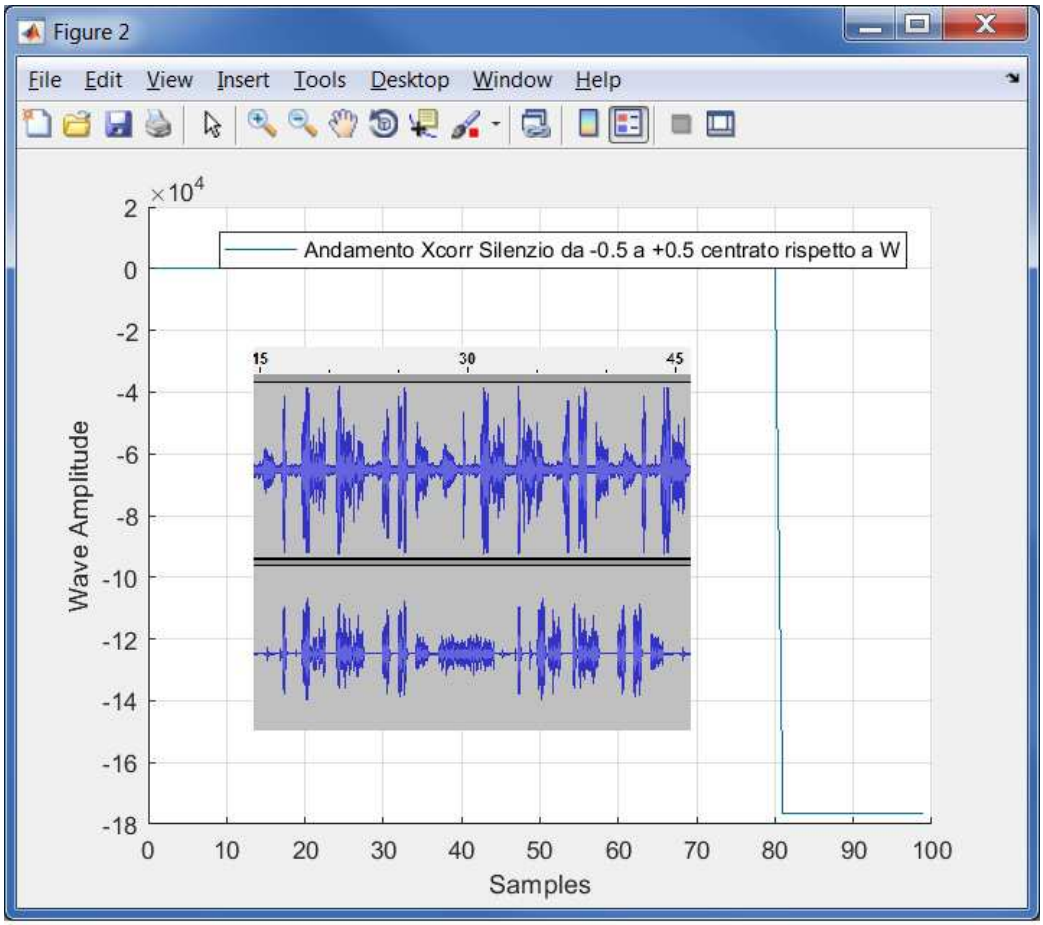


Figure 2
File Edit View Insert Tools Desktop Window Help
×10^4
Andamento Xcorr Silenzio da -0.5 a +0.5 centrato rispetto a W
Wave Amplitude
Samples

Si è poi verificato che al variare della dimensione della finestra di verifica o del segnale estraneo gli istanti di tempo relativi di commutazione dei valori, sempre solo 3, si spostavano rispetto ai punti caratteristici. Si sarebbe dovuto investigare sul tipo di relazione o sulla possibilità di mettere in relazione la posizione delle commutazioni all'ampiezza della finestra e del disturbo, per ricavarne una legge da utilizzare nel programma.

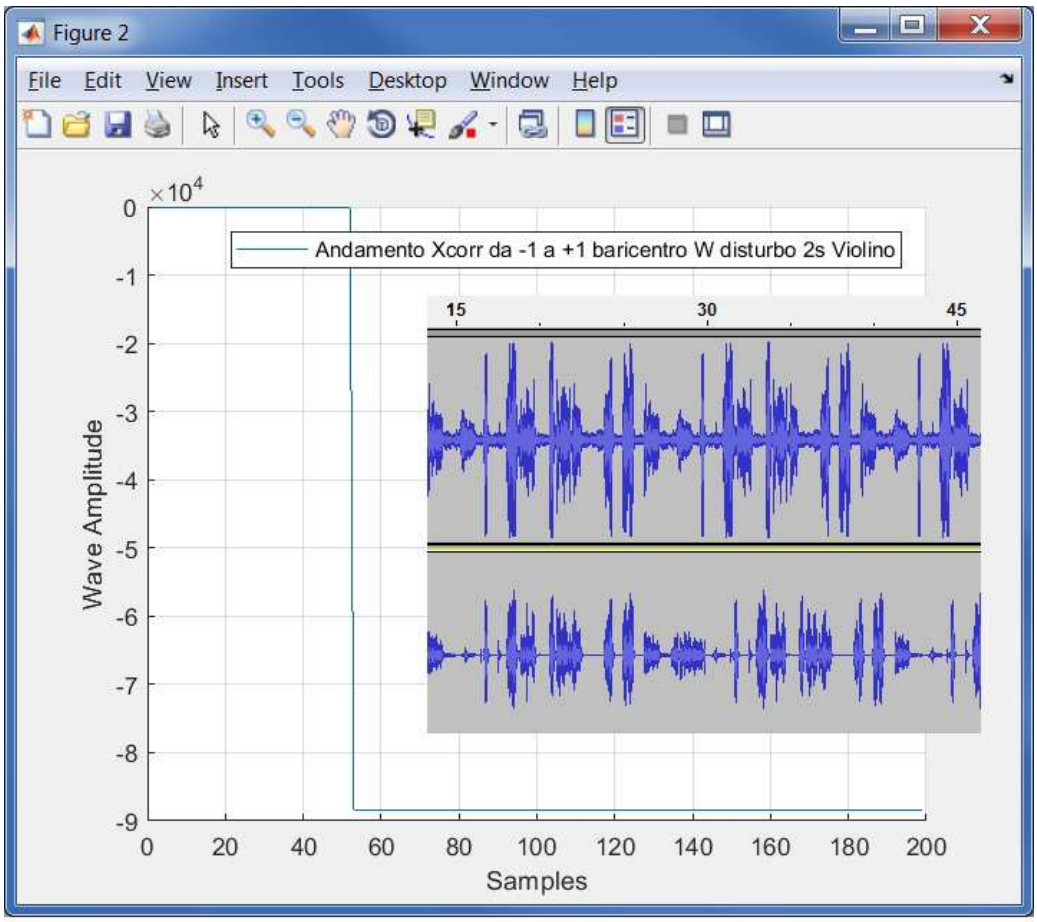


Figure 2
File Edit View Insert Tools Desktop Window Help
×10^4
Andamento Xcorr da -1 a +1 baricentro W disturbo 2s Violino
Wave Amplitude
Samples

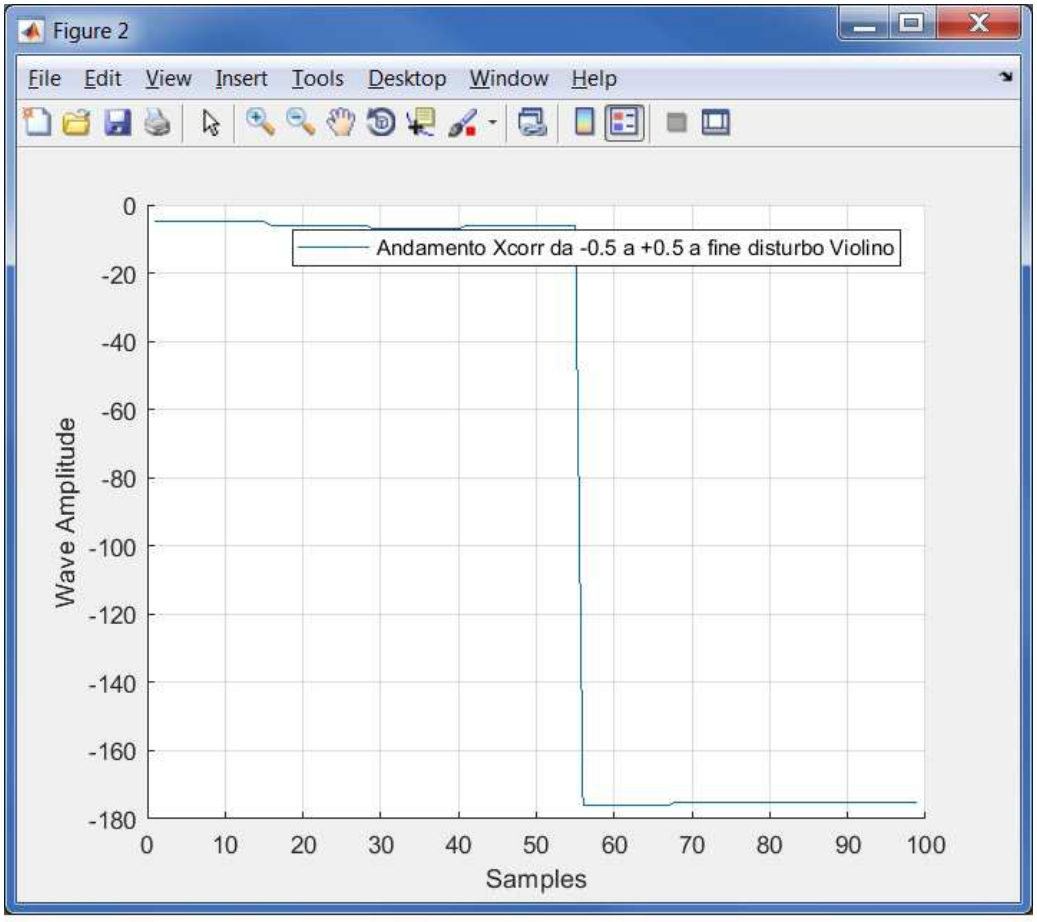


Figure 2
File Edit View Insert Tools Desktop Window Help
Andamento Xcorr da -0.5 a +0.5 a fine disturbo Violino
Wave Amplitude
Samples