

Analisi e Manipolazione Diretta di Formati Audio Compresi MP3 ed AAC

Dr. Giancarlo Vercellesi

LIM – Laboratorio di Informatica Musicale

DICo – Dipartimento di Informatica e Comunicazione

Università degli Studi di Milano

1. Introduzione al contesto

- Analisi di segnali audio musicali
(*MIR – Music Information Retrieval*)
- Manipolazione di segnali audio musicali
(*Editing Audio*)

2. Analisi e manipolazione tradizionale in dominio non compresso PCM e diretta in dominio compresso MPEG/Audio

3. Tecniche di analisi e manipolazione diretta di codifiche audio compresse MP3

Parte I

Introduzione al contesto

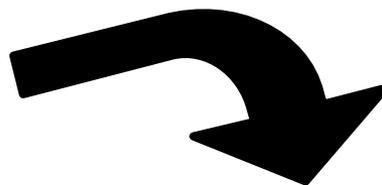
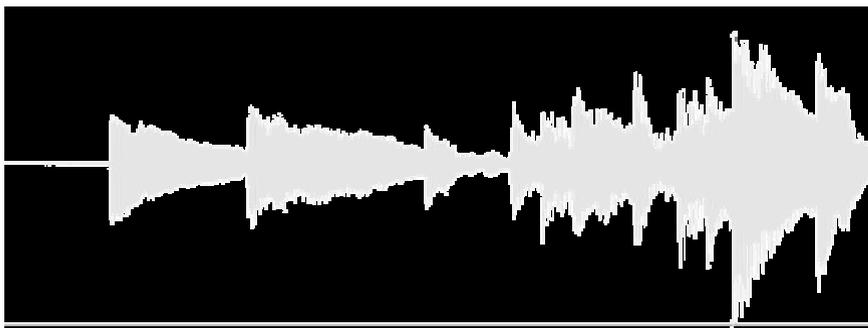
- Analisi di segnali audio musicali
- Manipolazione di segnali audio musicali

Analisi di segnali audio musicali

- Metodi e tecniche per il riconoscimento e l'estrazione automatica di informazione musicale significativa da segnali audio digitali
- Le informazioni estratte sono rappresentate in formati simbolico-musicali (XML)
 - MX ed MPEG-7 sono esempi di framework per la rappresentazione XML dell'informazione musicale

Dominio analisi automatica

Da segnale audio digitalizzato (compresso o non compresso) a partitura simbolica (XML)



A musical score for two hands in 3/4 time, showing the first five measures. The score is written in a single system with two staves. The top staff is labeled "1: LEFT HAND" and the bottom staff is labeled "2: RIGHT HAND". The key signature has one flat (B-flat) and the time signature is 3/4. The first measure of both hands is a whole rest. The second measure contains a quarter note G4 in the left hand and a quarter note G3 in the right hand. The third measure contains a quarter note F4 in the left hand and a quarter note F3 in the right hand. The fourth measure contains a quarter note E4 in the left hand and a quarter note E3 in the right hand. The fifth measure contains a quarter note D4 in the left hand and a quarter note D3 in the right hand. The score is numbered 1 through 5 at the top.

Operazioni di analisi automatiche

- Pitch tracking (audio level)
 - melody extraction (score level)
- Beat tracking (audio level)
 - Tempo extraction (score level)
 - Rhythm extraction (score level)
- Sincronizzazione automatica audio - partitura simbolica
 - DAN
- Trascrizione automatica partitura musicale (altezza, durate, timbro)
- Separazione sorgenti audio (de-mixing)
- Riconoscimento automatico genere, autore, modello interpretativo, ecc.

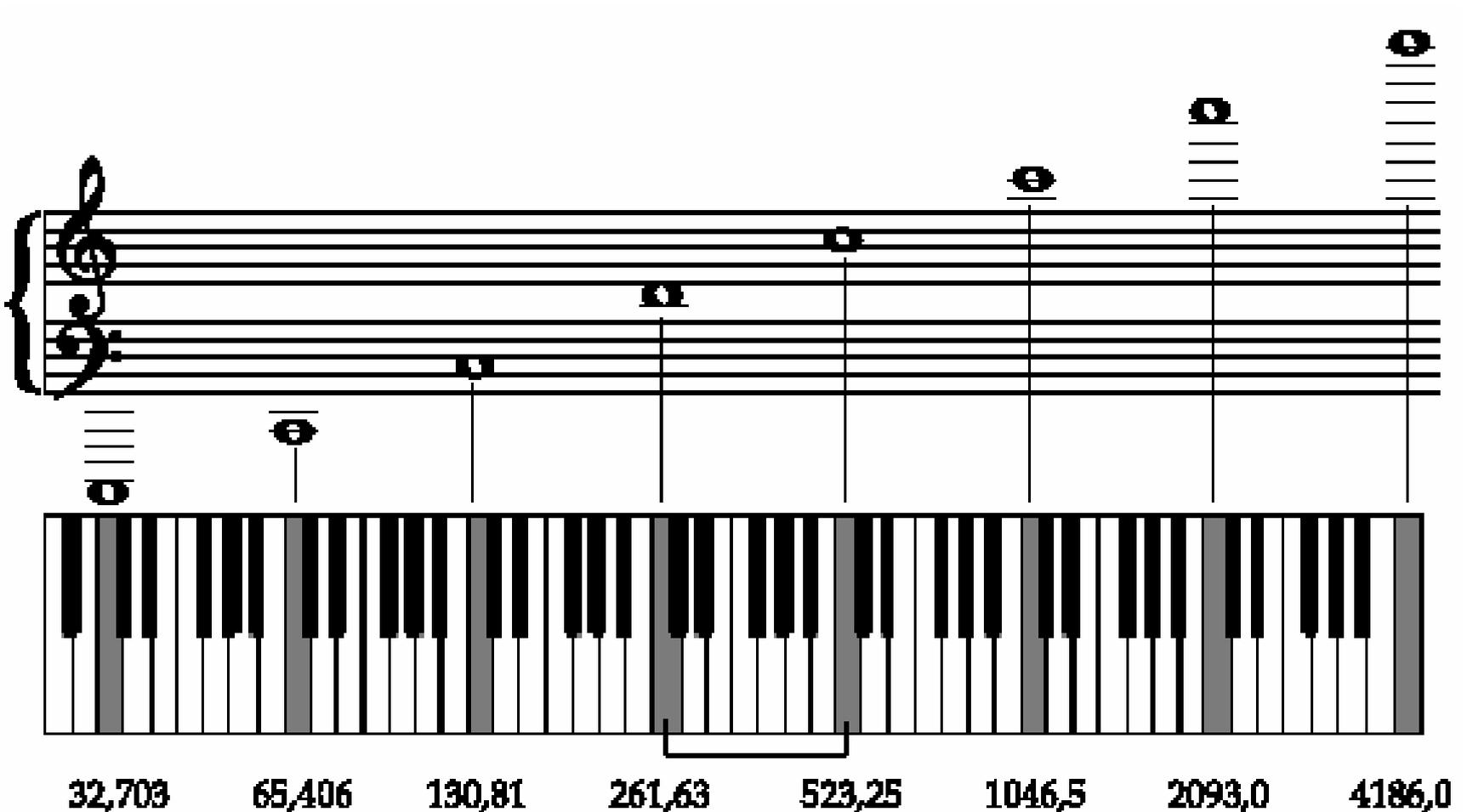
Beat Tracking

- Si definisce “beat” di un brano, una serie di fenomeni impulsivi dove il tempo musicale è ricavato dalla frequenza di questi impulsi. Questa definizione differenzia il beat dal ritmo, concetto che racchiude altre nozioni come gruppi, gerarchie e costruzioni ad alto livello.
- Un sistema di Beat Tracking deve essere in grado di individuare automaticamente la posizione di un Beat all’interno di una traccia audio

Pitch Tracking

- Sistema in grado di identificare automaticamente l'altezza di un suono in funzione del tessuto frequenziale in esso contenuto (fondamentale, armoniche e loro relazioni in ampiezza e frequenza)

Pitch Tracking



Sincronizzazione Automatica Audio-Partitura Simbolica

- Sistema in grado di trovare la relazione temporale tra audio e partitura per un qualsiasi segnale musicale in ingresso: lo scopo è quello di associare ad ogni nota presente in partitura la giusta posizione all'interno del segnale audio.
- Questo sistema è paragonabile a quello di un ascoltatore che, mano alla partitura, segue le note ascoltate nell'audio leggendole sul pentagramma

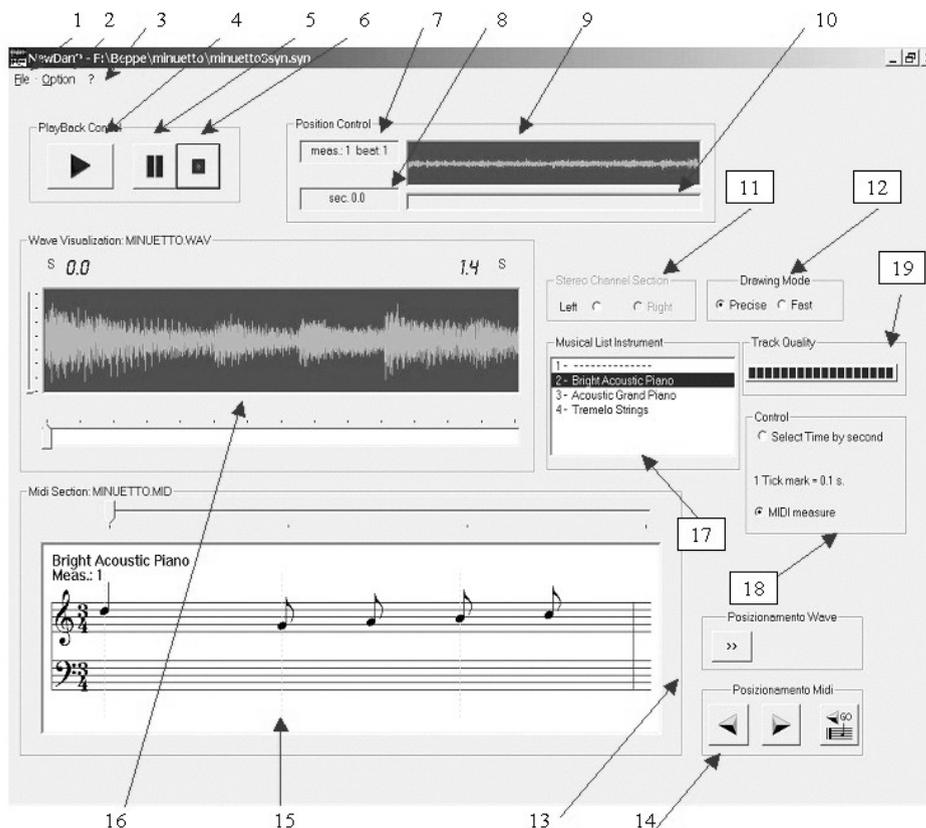
Applicazioni

- Navigazione sincronizzata tra segnali audio e partiture musicali
- Indicizzazione e ricerca di materiali audio eseguite ad un livello simbolico e non numerico: più semplici, efficaci e veloci
- Classificazione automatica di archivi musicali
- Error concealment: possibilità di rigenerare parte di un segnale audio corrotto in funzione della sua struttura simbolica, precedentemente estratta per via automatica
-

Tipologie segnali audio trattabili

- Monofonici e monotimbrici (% successo alta)
- Monofonici e politimbrici (medio-alta)
- Polifonici e monotimbrici (media)
- Polifonici e politimbrici (medio-bassa)

DAN – Digital Audio Navigator



<http://www.lim.dico.unimi.it/demo/dan/dan.htm>

Manipolazione di segnali audio musicali

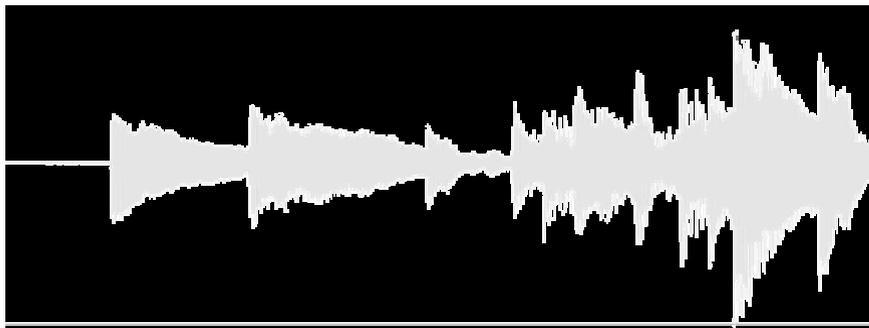
- Algoritmi numerici che permettono la modifica delle varie caratteristiche di segnali audio digitalizzati (editing audio)

Se applichiamo la funzione di manipolazione F ad $x(n)$, otteniamo un segnale manipolato $F(x(n))$ (con n intero razionale):

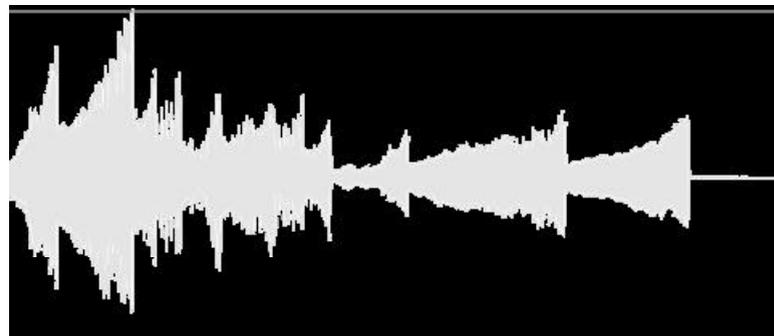
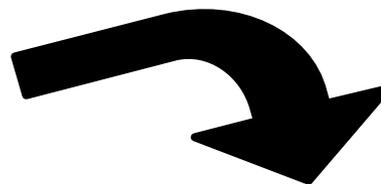
$$x(n) \rightarrow F \rightarrow F(x(n))$$

Dominio manipolazione

Da segnale audio a segnale audio (compresso/non compresso)



F = Time inversion



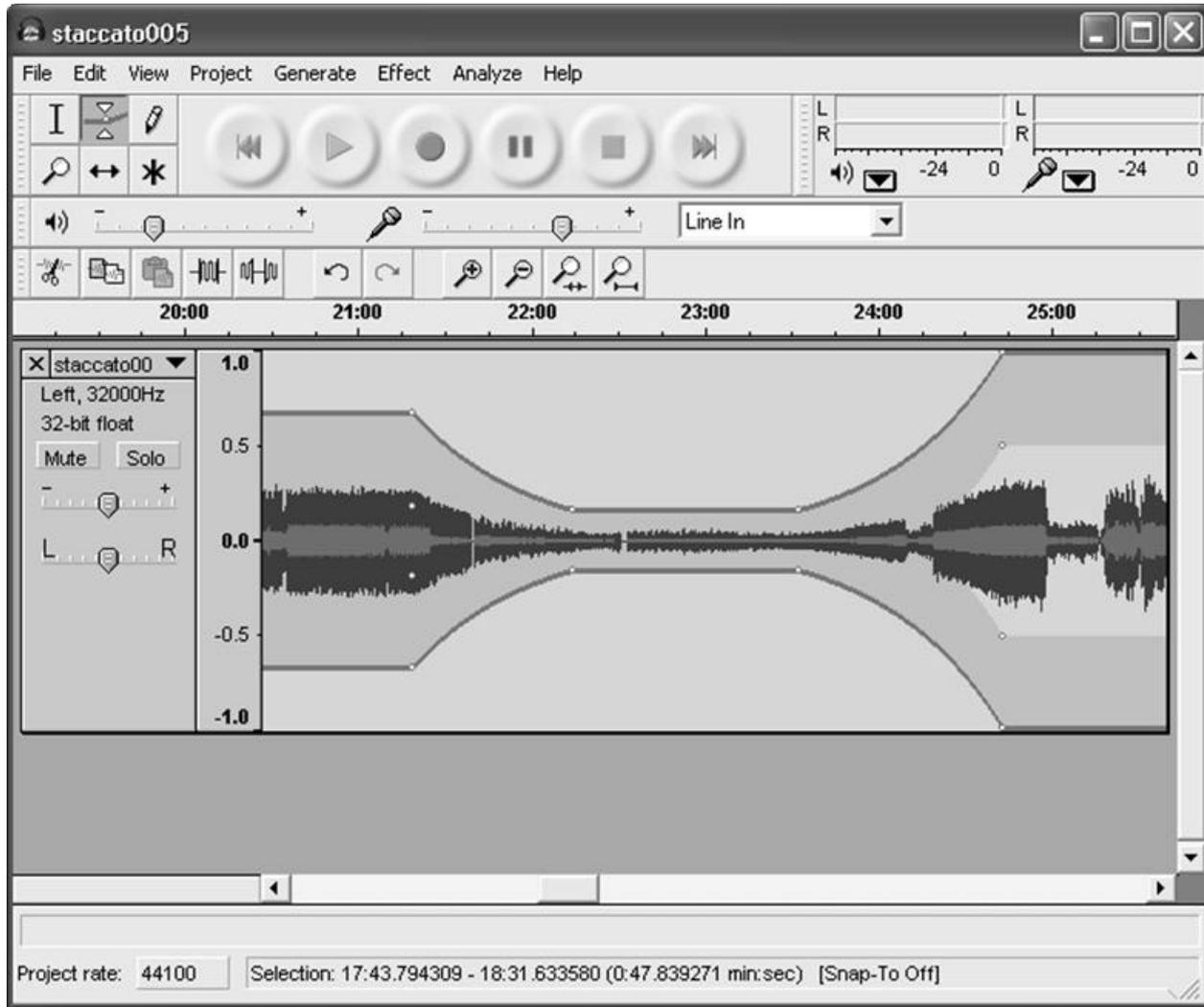
Operazioni di manipolazione

- Modifica della struttura temporale: taglia/copia/incolla, ecc.
- Modifica dei parametri del segnale: Fc, codifica di canale, risoluzione quantizzazione, sistema di codifica e rappresentazione, ecc.
- Modifica della dinamica: volume, normalizzazione, ecc.
- Modifica dello spettro: filtri, equalizzatori, ecc.
- Effetti: delay, echi, riverberi, effetti speciali (wah-wah), ecc.
- Time-frequency processing: time stretching, pitch/frequency shifting
-

Applicazioni

- Applicazioni consumer HW e SW
- Digitalizzazione e archiviazione
- Post-produzione audio - mastering
- Adattamento dei segnali a disturbi trasmissivi (streaming via cavo e via etere)
- Acustica architettonica
- Audiologia (protesi acustiche, ecc.)
-

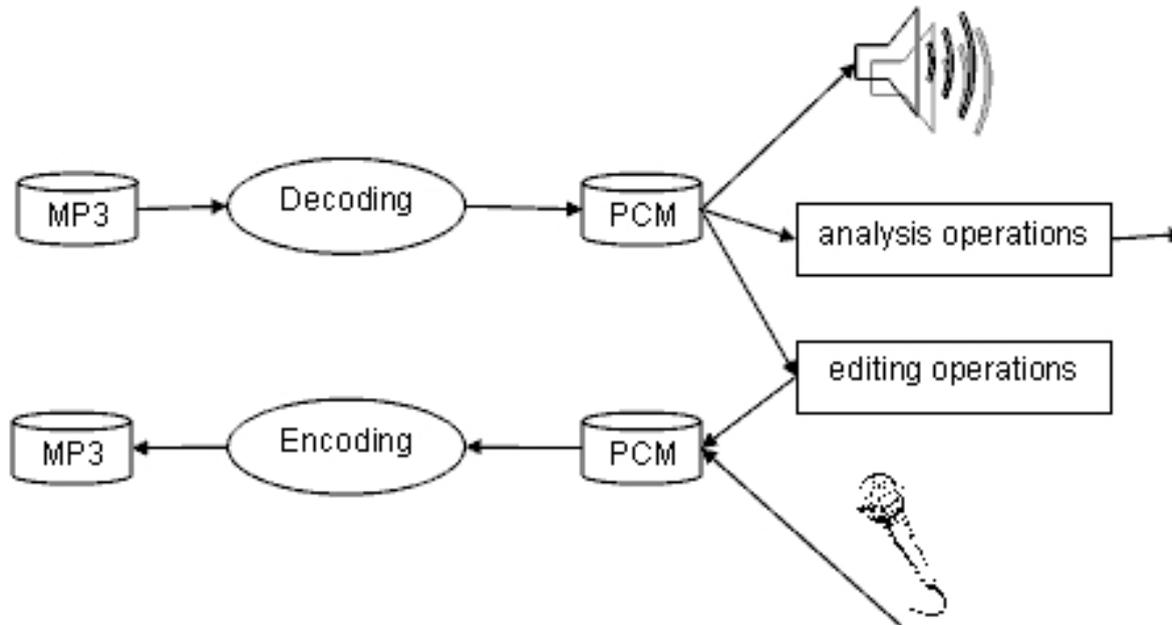
Audacity: esempio di editor audio



Parte II

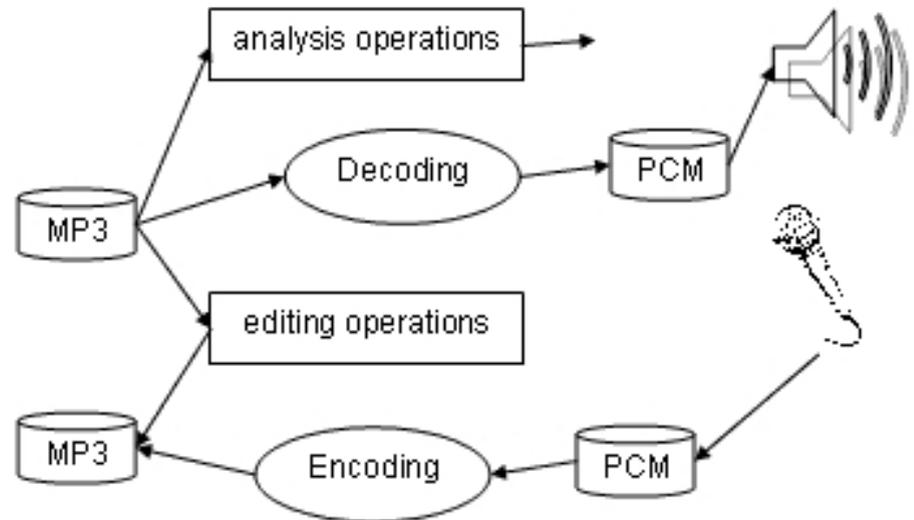
Analisi e manipolazione
tradizionale in dominio
non compresso PCM
e diretta in dominio
compresso MPEG/Audio

Analisi e manipolazione MP3 tradizionale (PCM)



Analisi e manipolazione diretta MPEG / Audio

- Manipolazioni eseguite direttamente sul formato compresso, decodificando ciò che è realmente necessario in base alla funzione da sviluppare
- Il decoder è necessario per la sola riproduzione del brano (i campioni PCM vanno spediti al DAC della scheda audio)
- L'encoder è necessario per la sola acquisizione audio da sorgenti esterne (i campioni PCM arrivano dall'ADC del sistema di input)



Principali differenze tra i due approcci

- Con l'approccio PCM
 - si opera nel dominio del tempo
 - l'informazione contenuta generalmente non ha subito tagli a livello di spettro del segnale
 - lo stato dell'arte fornisce una innumerevole quantità di algoritmi
- Con l'approccio MP3/AAC
 - si opera nel dominio delle frequenze
 - l'informazione contenuta può aver subito perdite informative (anche significative)
 - lo stato dell'arte fornisce poco o nulla
 - le tecniche di analisi e manipolazione devono essere sviluppate ad hoc in funzione del formato di codifica su cui si va ad operare

Esempio di editing audio PCM

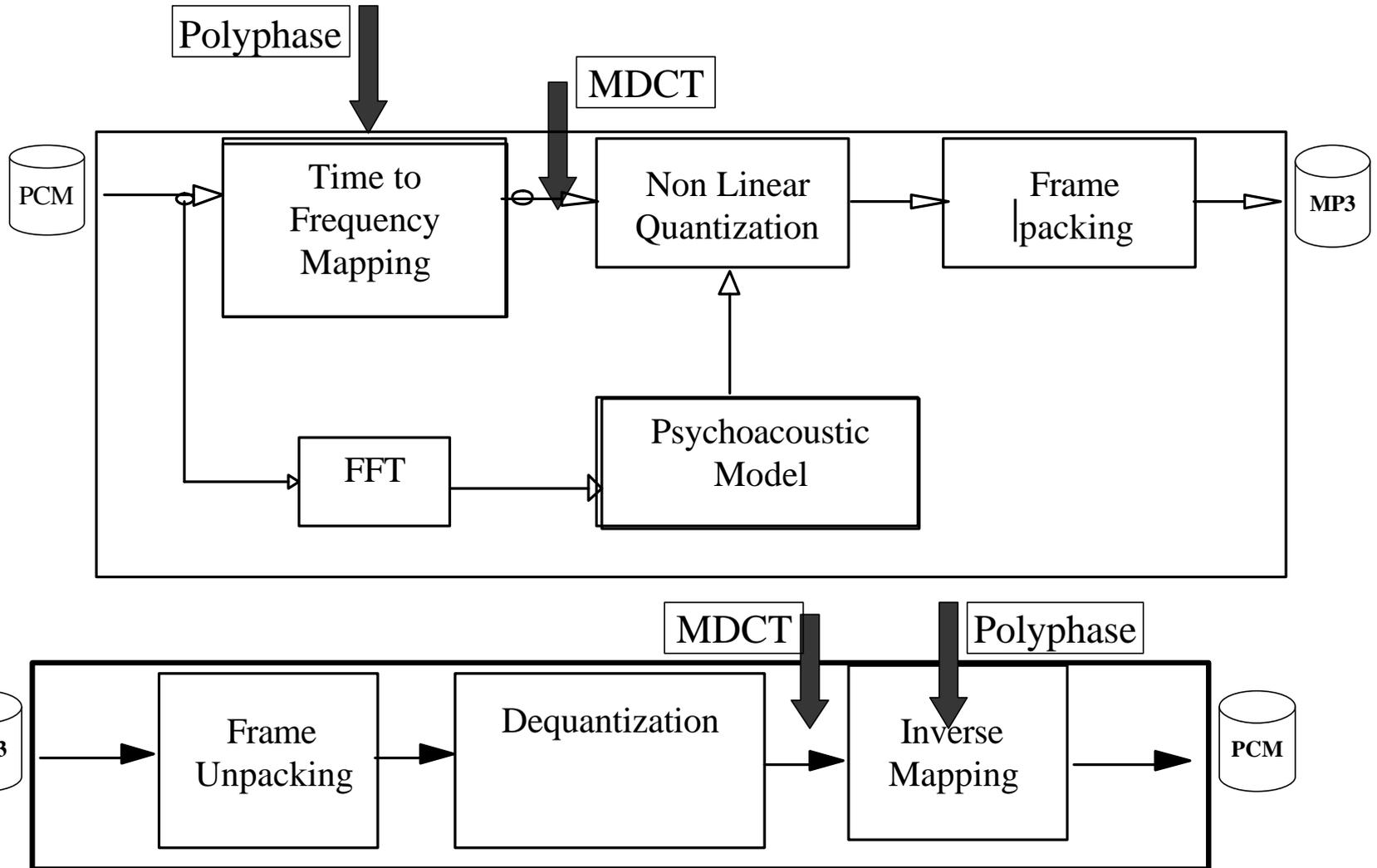
Pseudocodice per l'inversione temporale di un bitstream PCM

```
Void invert_wave(char *pathR, char *pathW)
{
    [y_pcm, Fs, nbit] = wavread(path);    //carico file wav
    [n_samples, n_channel] = size (y_pcm);

    for (ch=0; ch++; ch < n_channel)    //eseguo l'inversione
        for (i=0; i++; i < nr)
            y_pcm_inv[i][ch] = y_pcm[nsample - i][ch];

    wavwrite(y_pcm_inv, Fs, nbit, pathW) // salvo file wav invertito
}
```

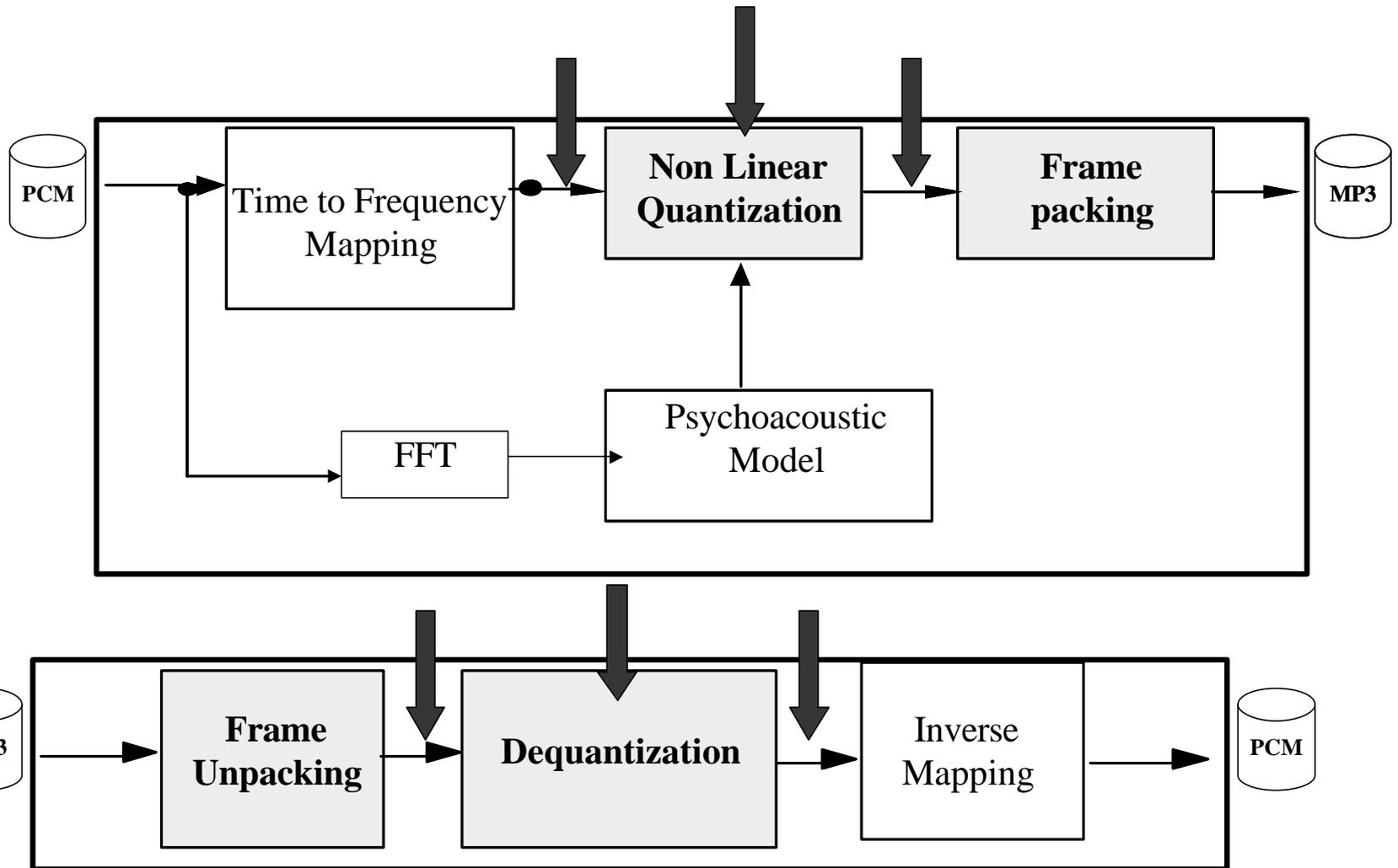
Analisi diretta MP3



Analisi: vantaggi e svantaggi

- Vantaggi
 - Risparmio di tempo
 - Risparmio di spazio
 - Possibile migliore qualità dei risultati
 - Patent Licence
- Svantaggi
 - Minore risoluzione temporale
 - Parametri di analisi spettrale fissati e non modificabili

Manipolazione diretta MP3



Manipolazione: vantaggi e svantaggi

- **Vantaggi**

- Risparmio di tempo
- Risparmio di spazio
- Mantenimento della qualità percepita (ITU tests)
- Patent Licence

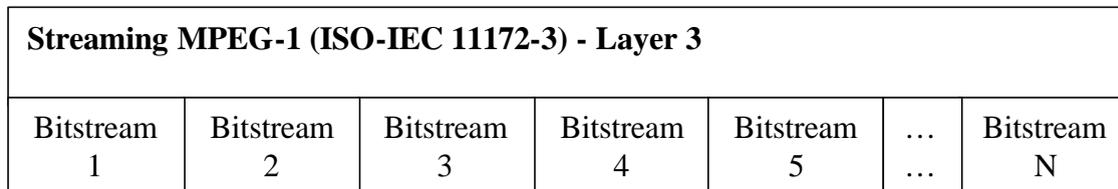
- **Svantaggi**

- Minore risoluzione temporale
- Parametri di analisi spettrale fissati e non modificabili

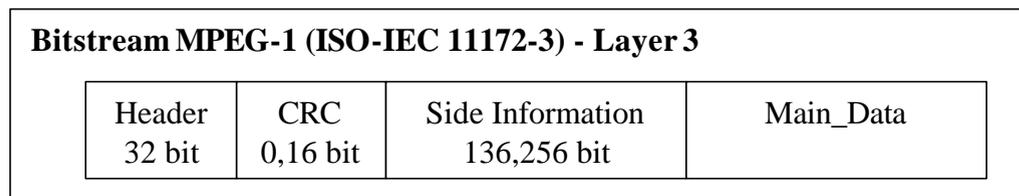
Cenni sul formato MP3

Formato MPEG Layer 3

- Sequenza di Frame / Bitstream audio decodificabili indipendentemente
- Ogni frame codifica 576 - 1152 sample PCM

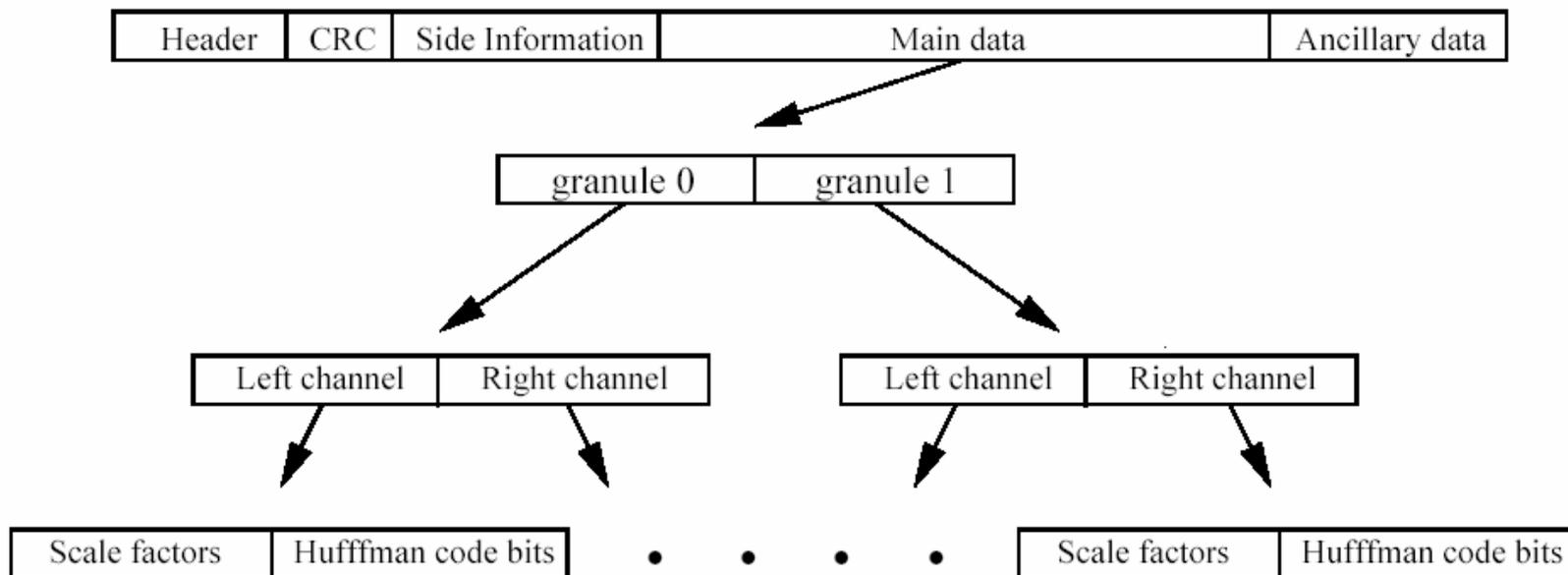


- “Header” e “Side Information” permettono di decodificare correttamente i dati audio contenuti nel “Main_Data”
- CRC (campo per la validazione dei dati) facoltativo



Formato MPEG Layer 3

- Main_Data suddivisi in dati audio veri e propri ed Ancillary Data
- I dati audio veri e propri sono suddivisi in due *granuli*, ognuno dei quali corrisponde a 576 sample PCM (in MPEG-2 ed MPEG-2.5 c'è un solo granulo)
 - ScaleFactor: pesi associati a gruppi di valori frequenziali, raggruppati secondo la scala di Bark (bande critiche). Questi valori servono a “colorare” il rumore di quantizzazione
 - HuffmanCodeBits: valori frequenziali quantizzati e codificati in Huffman
- Ancillary Data: contengono informazioni extra come i metadata, ecc.



Parte III

Metodi per l'analisi e la manipolazione diretta di codifiche audio MP3

Metodi per l'analisi diretta di codifiche audio MP3

Dominio MP3 considerato per l'analisi automatica

- L'analisi viene eseguita sui singoli granuli, corrispondenti a 576 campioni PCM.
 - Frame MPEG-1 composto da due granuli
 - Frame MPEG-2 ed MPEG-2.5 composto da un singolo granulo
- La risoluzione d'analisi diretta è dunque pari al granulo (equivalente al frame in MPEG-2 ed MPEG-2.5)

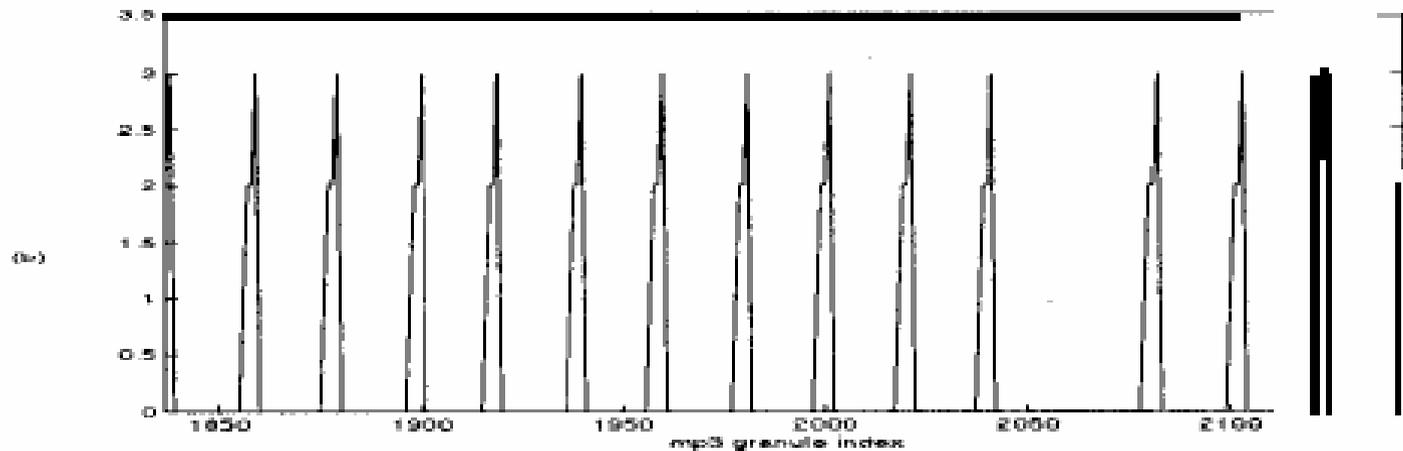
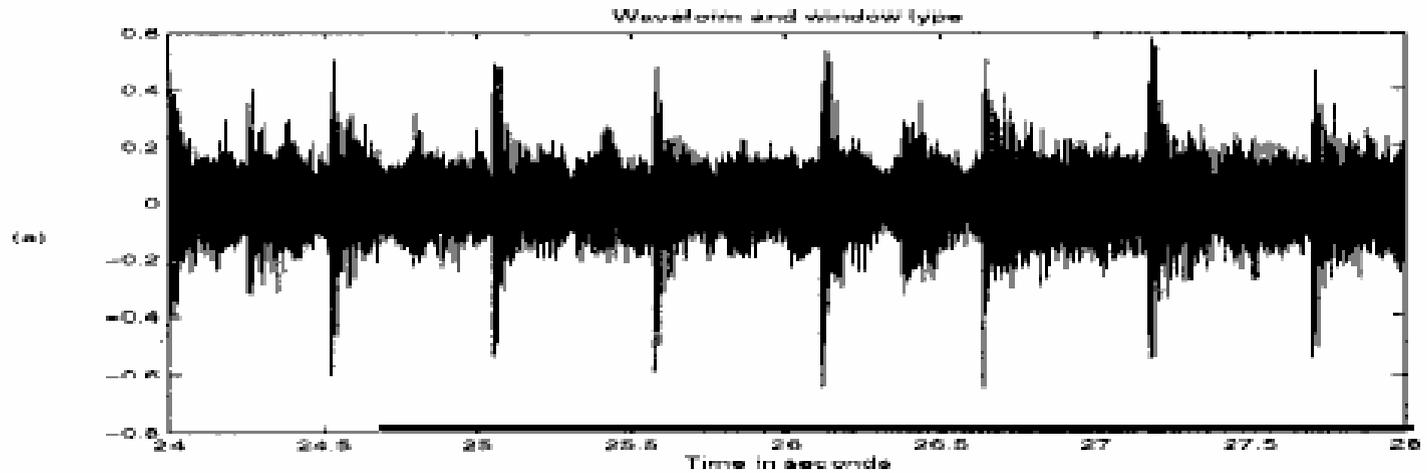
Coefficienti MDCT o filtro polifasico?

- E' possibile operare su entrambe le tipologie di rappresentazione
- A differenza di MP3, la codifica AAC non prevede l'uso di un banco di filtri polifasico. Le nostre ricerche (presso il LIM) si sono perciò principalmente orientate sui **coefficienti MDCT** in modo da poter sviluppare algoritmi compatibili, sia per codifiche MP3 che per codifiche AAC

Perché usare il WSP

- In fase di analisi spettrale, il codec MP3 può utilizzare finestre lunghe (maggiore risoluzione frequenziale) o corte (maggiore risoluzione temporale) in funzione del valore del parametro di *entropia psicoacustica* fornito dal modello psicoacustico.
- Tale accorgimento permette di evitare o ridurre gli effetti di pre-eco in fase di resintesi del segnale compresso

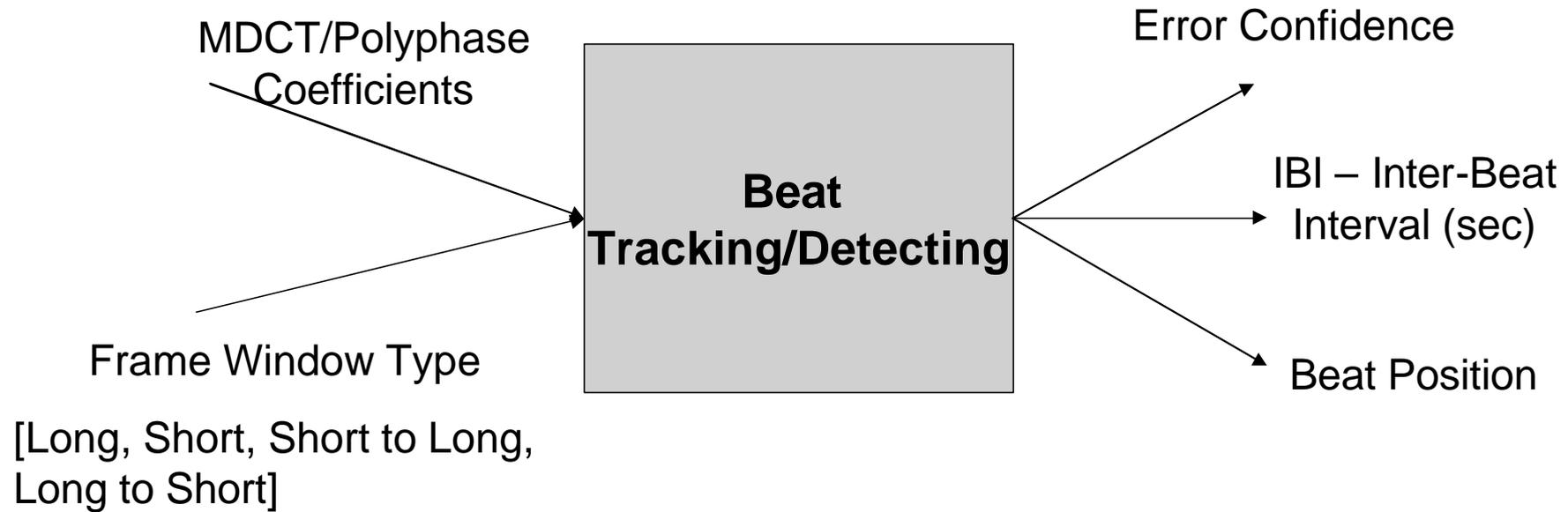
Perché usare il WSP



Considerazioni su WSP

- Il Window Switching Pattern dipende da:
 - Bontà dell'encoder
 - Valori di bitrate e frequenza di campionamento
 - Presenza della tecnica del Bit Reservoir
 - **Utente: molti codec permettono di creare MP3 senza *short block***
- Window Switching Pattern non è dunque un indicazione sempre impiegabile ed affidabile.

Beat Tracking



Algoritmo di Beat Tracking

Implementabile sfruttando due differenti metodologie:

- Windows Switching Pattern (WSP) Detection
- MDCT (o Polyphase) Detection

MDCT and WSP Detection

- MDCT Detector analizza l'energia presente nelle sottobande critiche e l'energia media totale di ogni singolo granulo.
- Se i valori di energia precedentemente determinati superano dei valori di soglia (valori predeterminati a priori o calcolati dinamicamente in funzione delle analisi passate), un beat -ed il relativo *Error Confidence*- è stato identificato.
- Se in un singolo granulo, risultano esserci più beat candidati, viene selezionato quello con minore *Error Confidence*.

MDCT and WSP Detection

- WSP viene semplicemente estratto dai campi del formato MP3
- L'output fornito da MDCT Detection è confrontato con il WSP (quando presente); se i due risultati differiscono lievemente, viene data priorità al beat identificato da WSP (generalmente molto affidabile e presente anche nei codec AAC).

Error Concealment e Beat Detection

- Le tecniche di Error Concealment hanno l'obiettivo di occultare opportunamente pacchetti audio persi durante trasmissioni di rete (es: streaming over IP)
- L'occultamento di questi pacchetti con altri ritmicamente simili -opportunamente identificati prima della trasmissione con Beat Tracking- fornirebbero segnali audio maggiormente piacevoli da un punto di vista percettivo.

Time Stretching e Beat Detection

- Le tecniche di Time Stretching permettono di allungare temporalmente un suono senza modificarne la natura spettrale.
- La ripetizione di frame ritmicamente simili –opportunamente identificati con Beat Tracking- porterebbe a segnali audio temporalmente allungati e percettivamente piacevoli

Pitch Tracking



[Long, Short, Short to Long,
Long to Short]

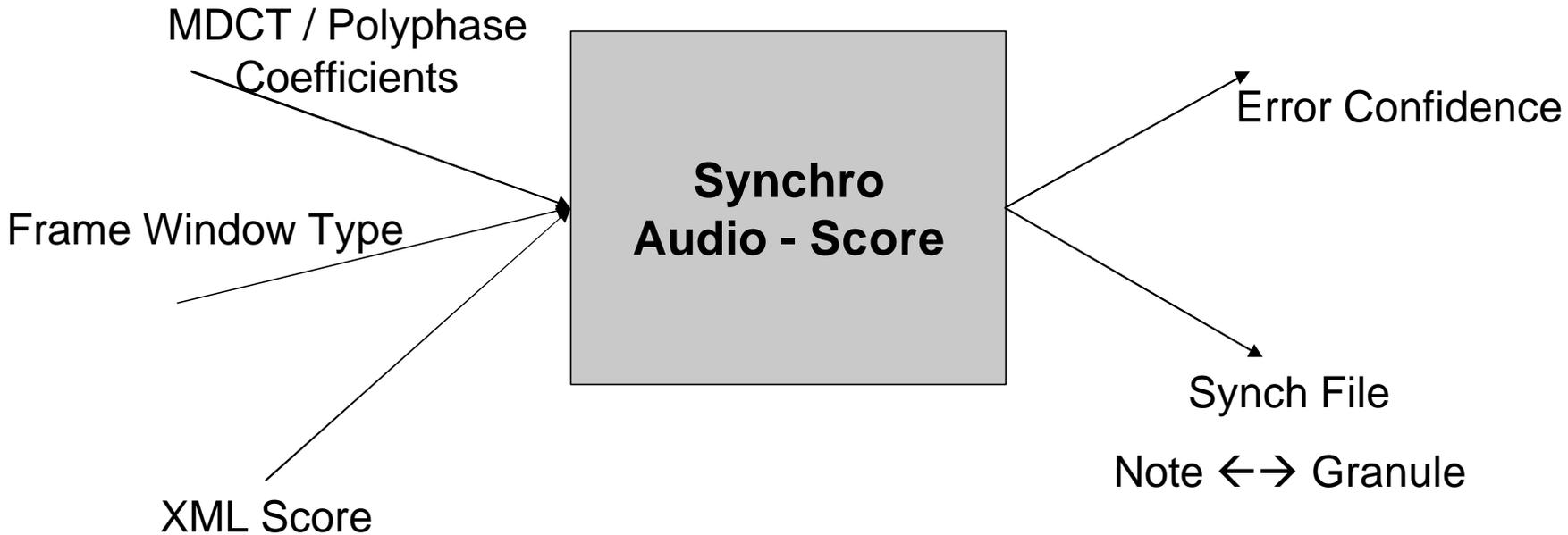
Algoritmo di Pitch Tracking

- Implementabile sfruttando il Windows Switching Pattern (WSP) ed i coefficienti MDCT o Polyphase
- WSP (se presente) utilizzato per identificare forti attacchi di nota e percussioni

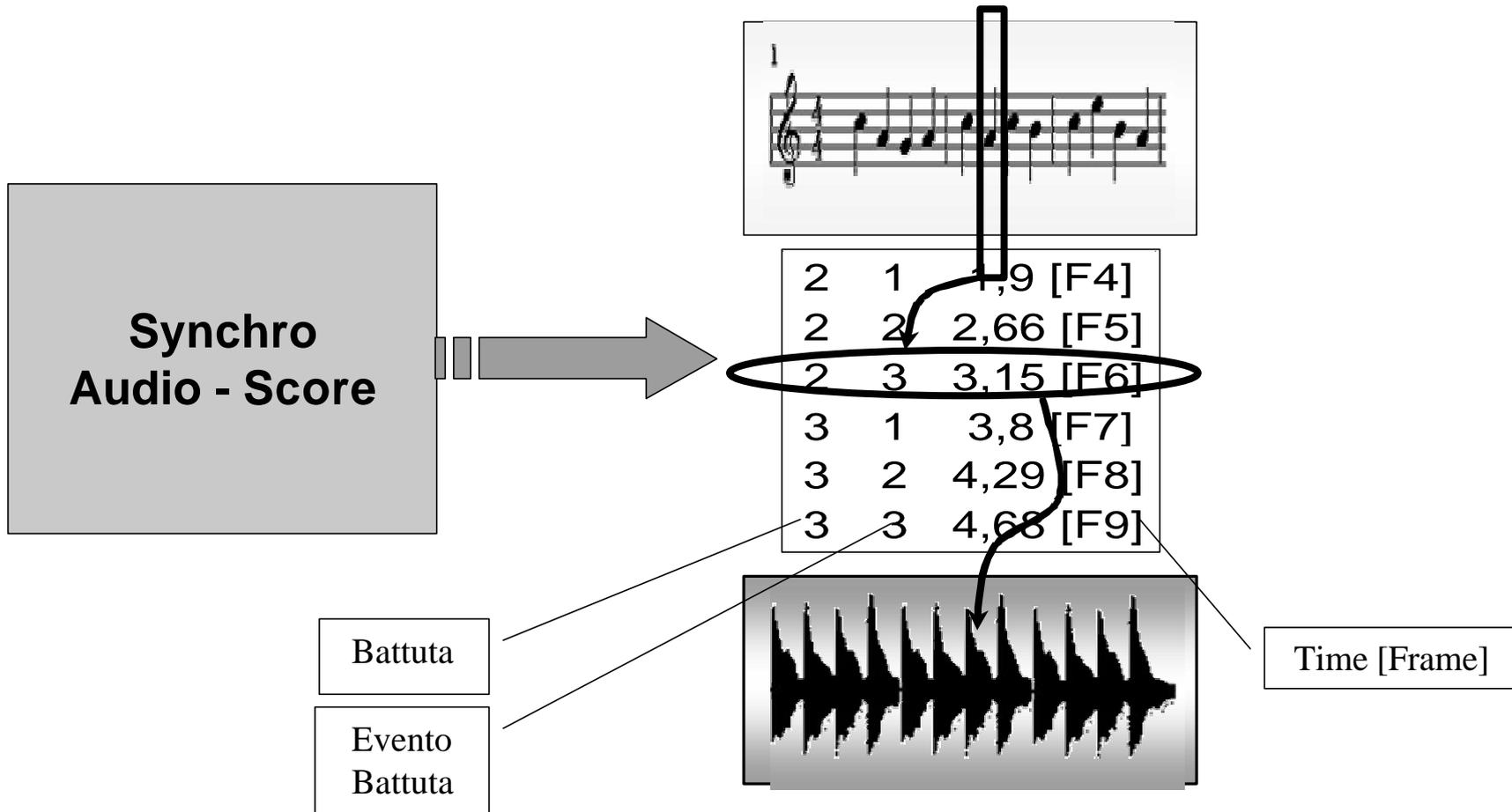
Algoritmo di Pitch Tracking

- Coefficienti del banco di filtri ibrido (MDCT o polyphase) usati per il calcolo dell'energia spettrale del granulo considerato.
- Dall'analisi della struttura energetica dello spettro del segnale si ricava la corrispondente altezza presente (e l'eventuale *error confidence*)
- Il calcolo dell'energia spettrale può essere effettuata:
 - nelle sottobande critiche
 - nelle 32 sottobande del banco di filtri polifasico
 - in opportunamente sottobande definite a priori, tali da facilitare il riconoscimento delle eventuali armoniche presenti (2F, 3F, ecc.)

Sincronizzazione Automatica Audio - Partitura Simbolica



Sincronizzazione Automatica Audio - Partitura Simbolica



Sincronizzazione Automatica Audio - Partitura Simbolica

- Implementabile sfruttando:
 - rappresentazione simbolico - esecutiva della partitura in XML
 - Windows Switching Pattern (WSP) e coefficienti MDCT o Polyphase di MP3
- WSP (se presente) utilizzato per identificare eventi di *note on* (attacchi di nota, percussioni)
- La sincronizzazione avviene tra singole note e granuli MP3

Differenti approcci per la sincronizzazione

- *Dall'audio alla partitura*: analisi del tessuto frequenziale effettuata sul segnale e successivamente confrontata a livello simbolico
- *Dalla partitura all'audio*: analisi del tessuto frequenziale effettuata sulla partitura e successivamente confrontata a livello audio

Differenti approcci per la sincronizzazione

- *Analisi fine del tessuto frequenziale:*
l'analisi è volta ad estrarre la fine struttura armonica contenuta in ogni singolo granulo
- *Analisi delle differenze nel tessuto frequenziale:* l'analisi rileva le differenze significative tra granuli adiacenti

Metodi per la manipolazione diretta di codifiche audio MP3

Prototipi SW

- **Taglia & Incolla - MP3 Editor 1.1**

- prototipo software, con interfaccia grafica e funzioni di *player*, che permette di eseguire le operazioni di editing:
 - Taglia / Copia / Incolla di frame MP3
 - Eliminazione di frame MP3
 - Inversione di frame MP3

- **Volume e Manipolazione Spettro - MP3 Direct EQ**

- prototipo software, con interfaccia grafica e funzioni di *player*, che permette di eseguire le operazioni di:
 - Modifica del volume
 - Filtri LP, HP, BP, BR
 - Equalizzatore a 10 bande

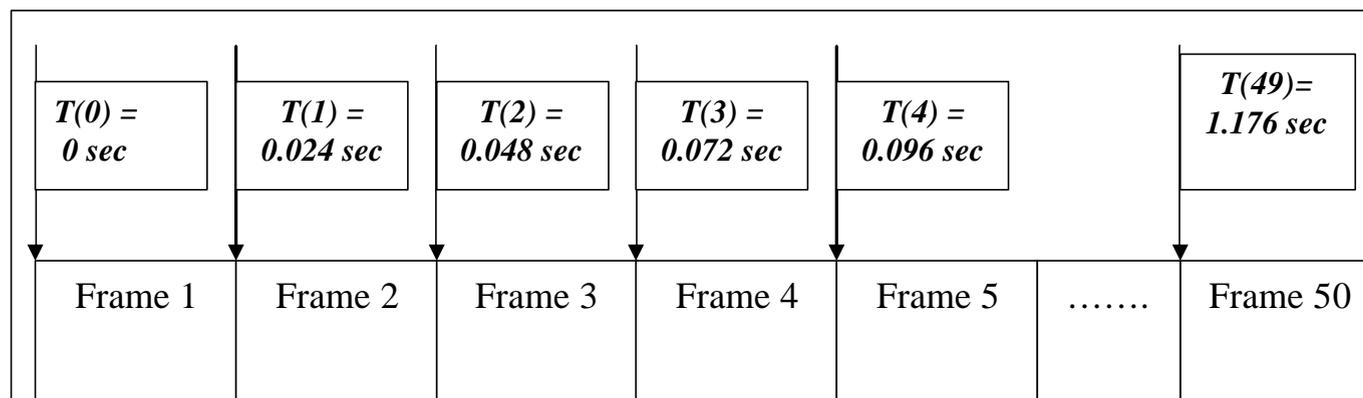
Download prototipi SW: www.lim.dico.unimi.it/demo/mp3directedit

Tecniche di Manipolazione Diretta

- Spostamento diretto di frame MP3
- Manipolazione dello spettro
- MP3 Bitrate Downgrading

Risoluzione Temporale

- La risoluzione temporale nell'editing diretto MP3 è pari alla dimensione di un frame
- Tale risoluzione è minore di quella PCM in quanto:
 - $[durata_frame_MPEG-1 = 1152 * T_c] > T_c$
 - $[durata_frame_MPEG-2_2.5 = 576 * T_c] > T_c$

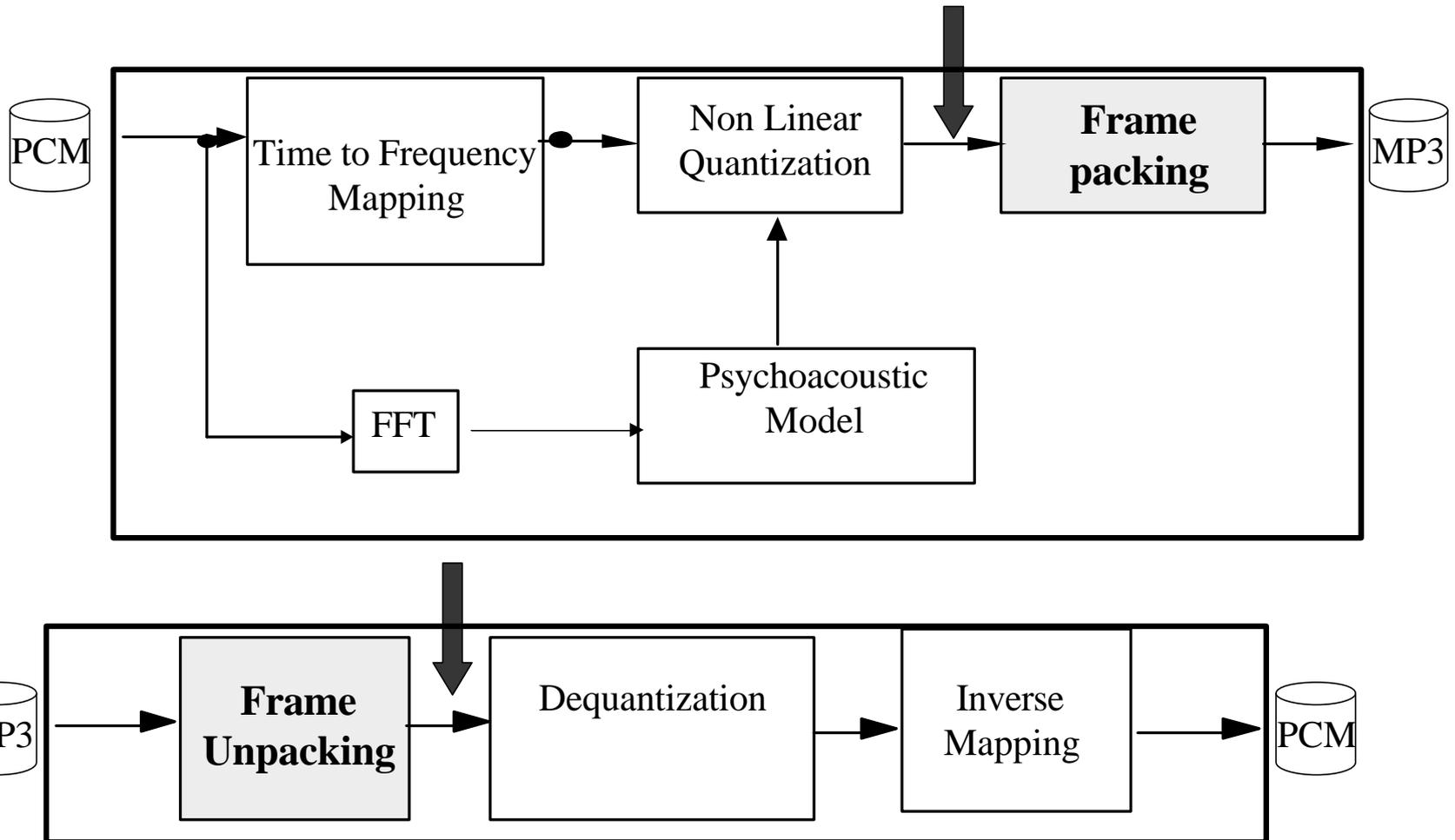


MPEG-1, $F_c = 48 \text{ KHz}$

Spostamento di frame MP3

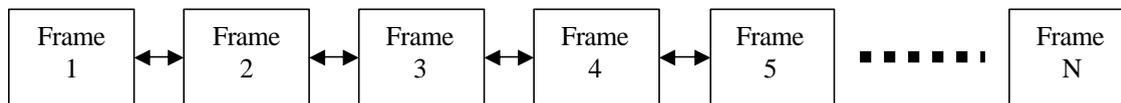
- Taglia / Copia / Incolla
- Eliminazione
- Inversione

“Spostamento di frame MP3”

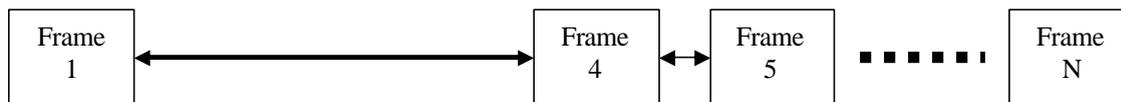


Spostamento di frame MP3

L'idea è quella di agire su una sequenza di frame indipendenti tra loro, potendo così spostarli all'interno di streaming MP3 senza problemi di decodifica.



Elimina F2 - F3

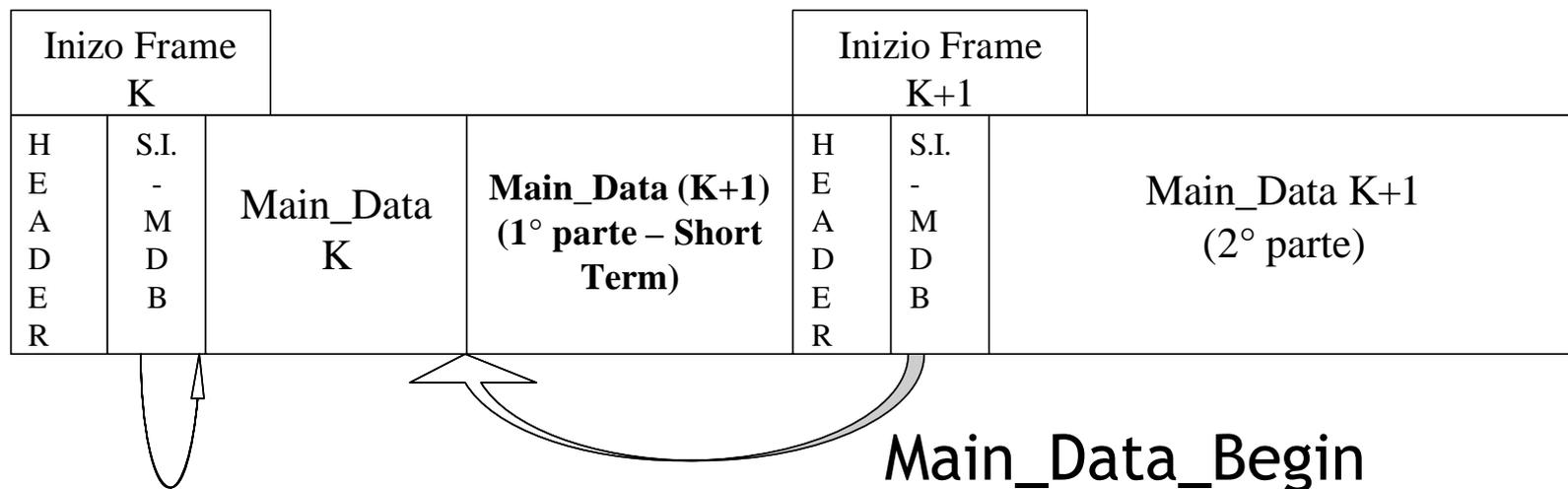


Problema del Bit Reservoir

- In fase di decodifica ogni frame è decodificabile in modo indipendente in quanto provvisto di proprio header.
- C'è però una dipendenza fisica tra frame MP3 che ne impedisce il loro spostamento diretto. Tale dipendenza è dovuta dalla presenza dell'uso della tecnica del Bit Reservoir

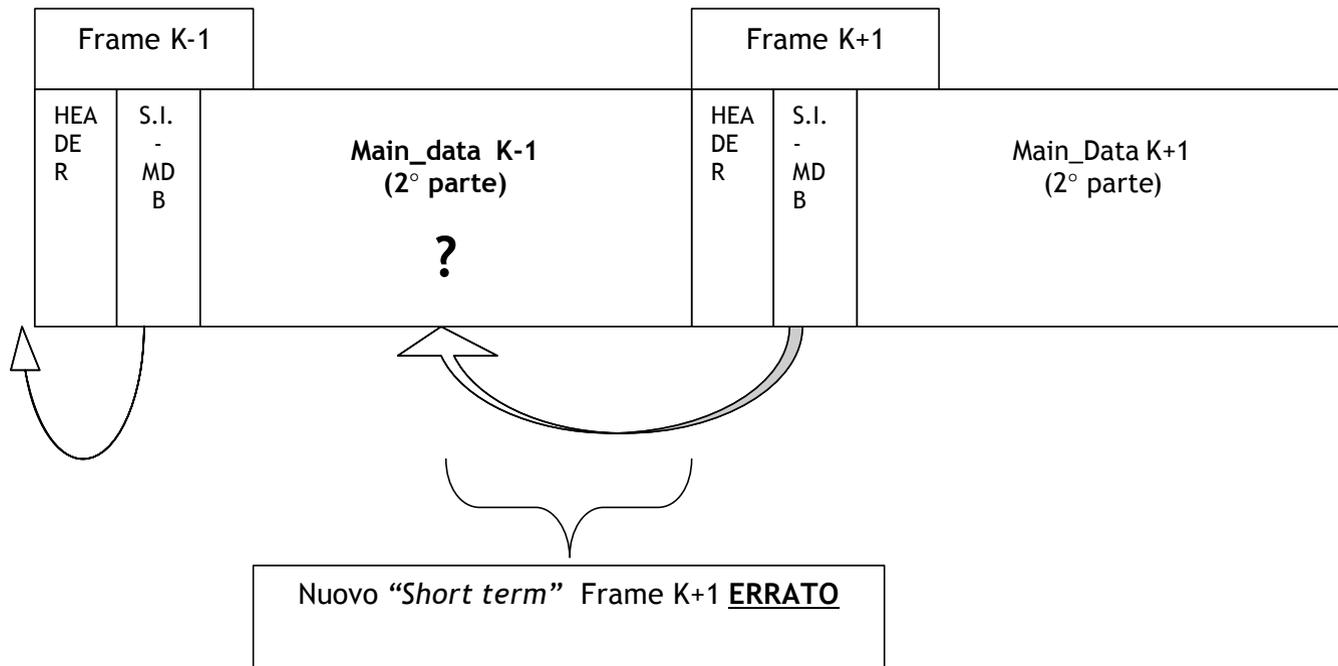
Problema del Bit Reservoir

- A causa della tecnica Bit Reservoir, i dati audio (main_data) non iniziano necessariamente nel proprio bitstream di appartenenza → frame tra loro interdipendenti



Problema del Bit Reservoir

- Se eliminassi il frame K avrei un'incongruenza nei main_data del frame K+1.
- Il frame K+1 non è più decodificabile correttamente



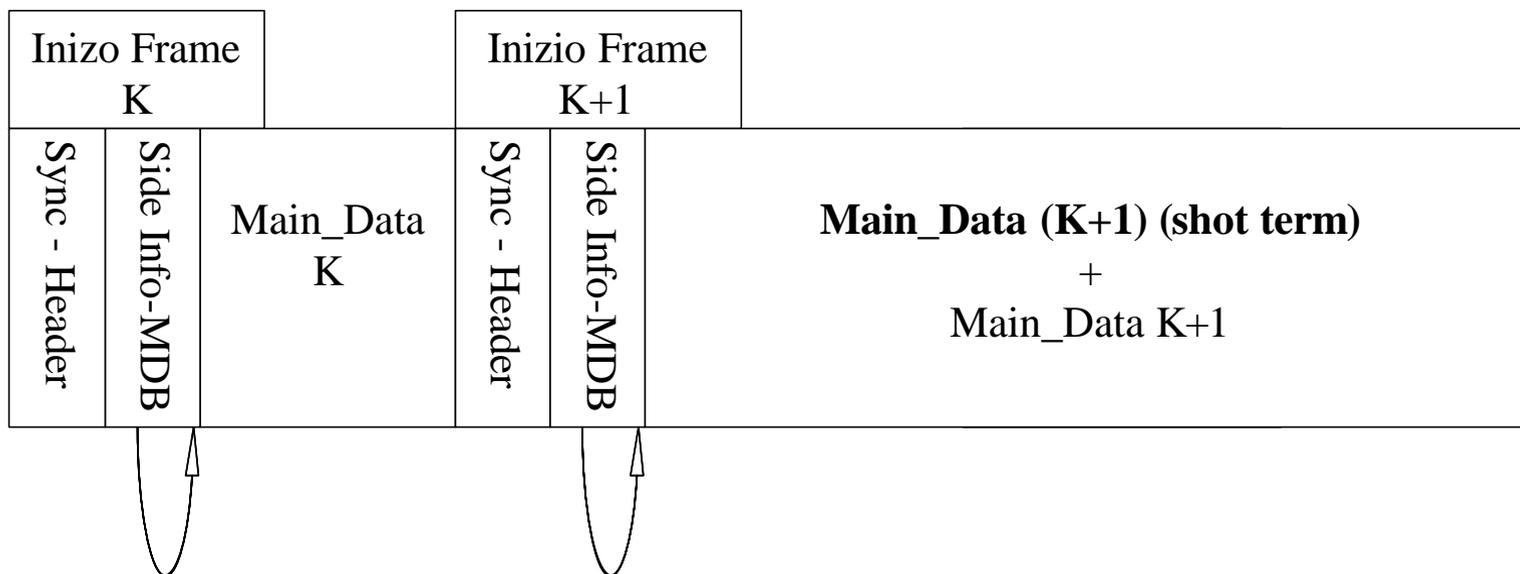
Tecnica dell'Allineamento

- E' necessario allineare i “main_data” con i propri frame rendendoli indipendenti l'uno con l'altro → **Tecnica dell'Allineamento**
- Operativamente si vanno a modificare opportunamente i valori del campo *bitrate* e *main_data_begin*

Tecnica dell'Allineamento

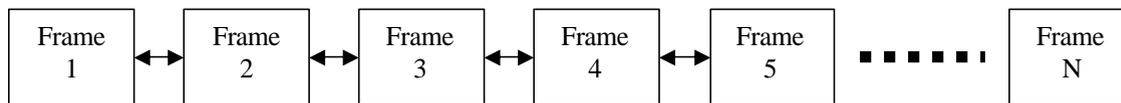
- FASE 1 – Loading and Decode: reperimento Frame - decodifica Header e Side Information - caricamento Main_Data nel loro bitstream
- FASE 2 – “Bit Reservoir” Suppression: soppressione della tecnica del “Bit Reservoir” ponendo a 0 il valore del campo `main_data_begin` di ogni frame
- FASE 3 – Average Bitrate: modifica dei bitrate affinché la dimensione dei bitstream sia minima e tale da contenere i propri *main_data*

Tecnica dell'Allineamento

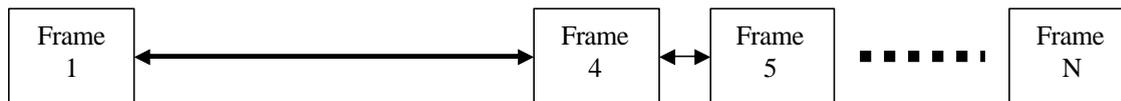


Funzioni di Editing

- Frame caricati in memoria come lista doppia ordinata
- Grazie a totale indipendenza tra i vari frame, è sufficiente spostare i singoli elementi della lista (frame) in modo da ottenere la sequenza audio voluta dall'utente
 - Taglia / copia / incolla; eliminazione; inversione

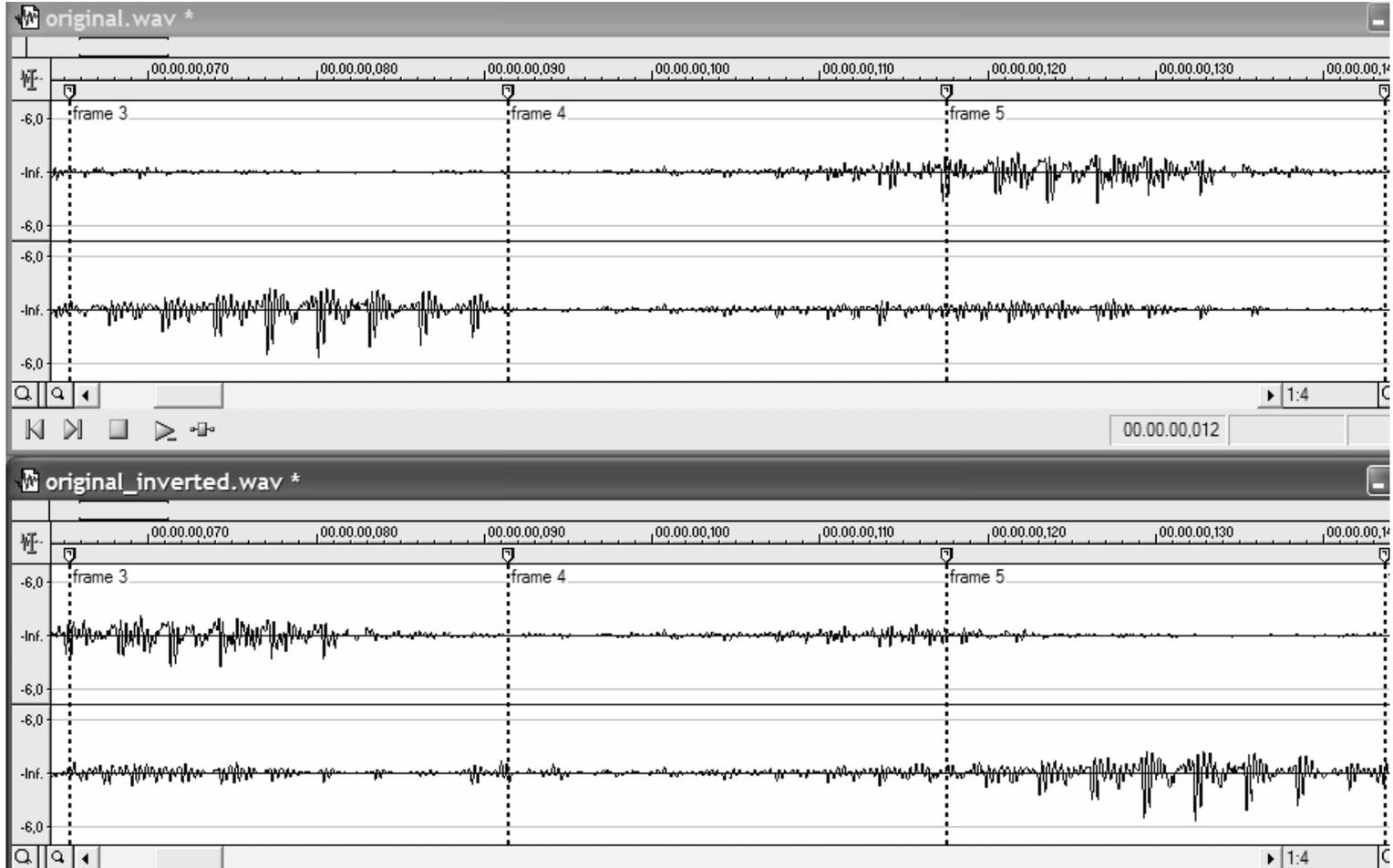


Elimina F2 - F3



Inversione di Frame MP3

Si inverte la sequenza temporale dei frame, non il loro contenuto.



Analisi Bitrate Medio

- Dopo allineamento si ottiene uno streaming MP3 con bitrate variabile dove:

$$\begin{array}{l} \text{bitrate medio} < \text{nuovo} \\ \text{originale} = \text{bitrate medio} < \text{Succ}(\text{bitrate medio originale}) \end{array}$$

Bitrate value

32 kbit/s

40 kbit/s

48 kbit/s

56 kbit/s

64 kbit/s

80 kbit/s

96 kbit/s

112 kbit/s

128 kbit/s

160 kbit/s

192 kbit/s

224 kbit/s

256 kbit/s

320 kbit/s

$\text{Succ}(\text{bitrate medio originale})$ = “bitrate successivo rispetto ai valori forniti da standard ISO (in tabella quelli previsti per MPEG-1 - ISO/IEC 11172-3)

ESEMPIO: MP3 con CBR a 128 Kbit/Sec

Dopo allineamento:

$128 \leq \text{“nuovo bitrate medio”} < 160$

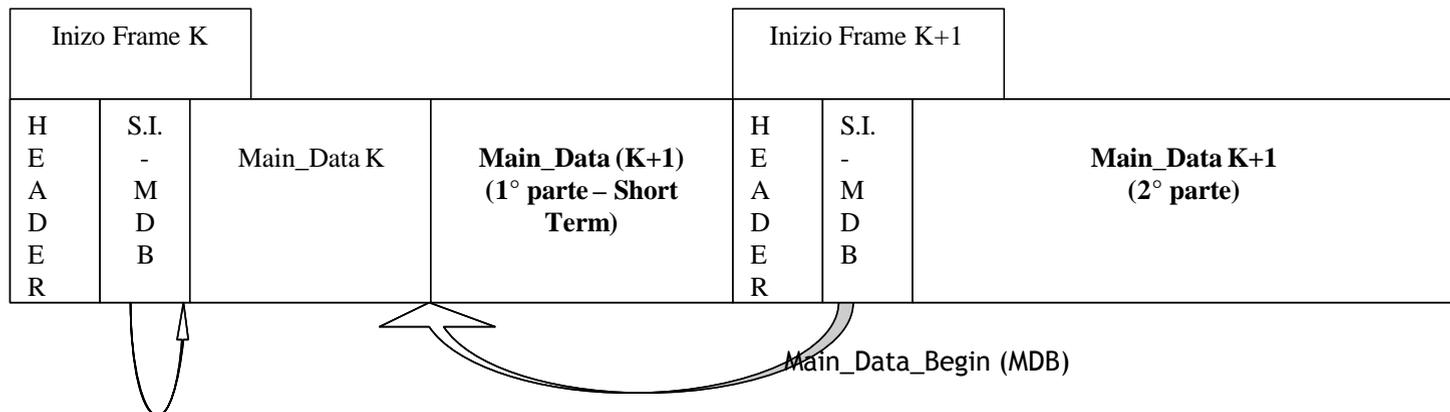
Analisi Bitrate Medio

- Se si volesse ottenere uno streaming MP3 con bitrate fisso (CBR) mantenendo l'allineamento tra frame, il bitrate dovrà essere pari al massimo valore presente.
- Esempio: streaming MP3 con 3 frame allineati e valori di bitrate 128, 160, 192. Per salvare l'MP3 allineato occorre un CBR a 192 Kbit/sec

Limiti Editing Diretto con MP3 non ISO-Compliant

- Lo standard ISO-IEC 11172-3 (MPEG-1) impone come massima dimensione del buffer di lettura di un decoder il valore 960 byte; questo valore equivale alla dimensione di un frame a 320 Kbit/sec ed $F_c = 48$ KHz
- In fase di encoding, il valore di `Main_Data_Begin` deve essere definito in modo tale che la dimensione del frame non superari questo valore massimo; se ciò non accade si ha un MP3 non ISO-Compliant non allineabile per l'editing diretto.

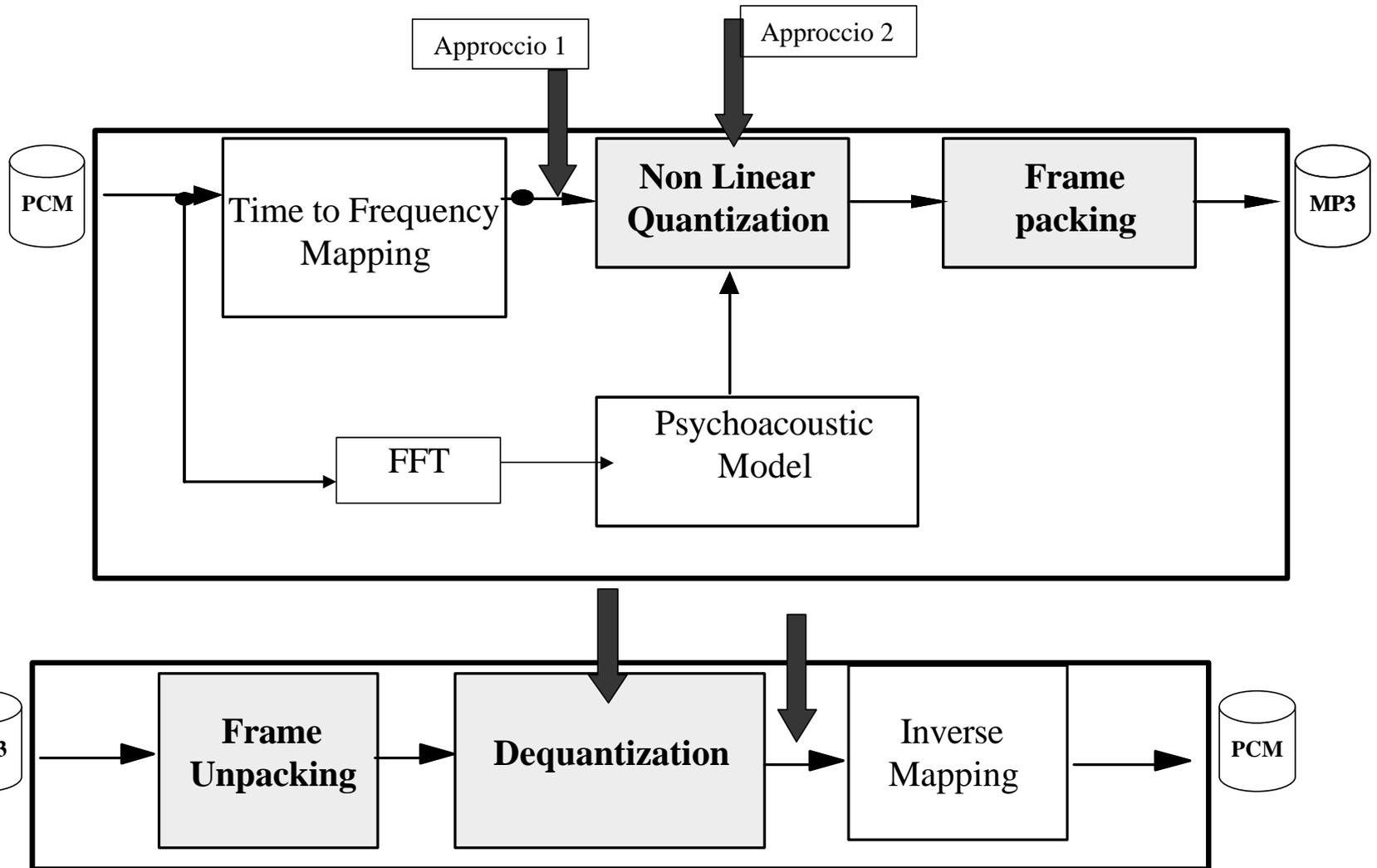
$$\text{Size (Frame K+1)} = \text{Size (Header e S.I.)} + \text{MDB} + \\ + \text{Size(Main_Data 2^\circ \text{ parte)} \leq \text{Size_frame_320kbit_noBitRes}$$



Manipolazione dello spettro

- Volume / Normalizzazione
- Filtri / EQ

Manipolazione diretta MP3



Formula di Dequantizzazione (Long block)

$$\mathbf{xr}[i] = \mathbf{sign}(\mathbf{is}[i]) * (|\mathbf{is}[i]|)^{(4/3)} * 2^{(1/4 * (\mathbf{GG}[\mathbf{gr}] - 210))} * \\ * 2^{(-\mathbf{SF_multiplier} * (\mathbf{SF}[\mathbf{sfb}][\mathbf{ch}][\mathbf{gr}] + \mathbf{preflag}[\mathbf{gr}] * \mathbf{pretab}[\mathbf{gr}]))}$$

dove:

- $\mathbf{xr}[576]$ = vettore frequenze non quantizzate di 576 elementi
- $\mathbf{is}[576]$ = vettore frequenze quantizzate di 576 elementi
- $\mathbf{GG}[\mathbf{gr}]$ = Global_Gain del granulo “gr”
- $\mathbf{SF}[\mathbf{ch}][\mathbf{gr}]$ = ScaleFactor del granulo “gr” e canale “ch”
- $\mathbf{SF_multiplier}$ = ScaleFactor Multiplier
- $\mathbf{preflag}[\mathbf{gr}] * \mathbf{pretab}[\mathbf{gr}]$: offset da aggiungere al valore di SF se è attivato il flag di preenfasi nell’header MP3

OSS su $|\mathbf{xr}[i]|^{(3/4)}$: in fase di encoding, la dinamica totale del segnale viene ridotta di $3/4 \rightarrow$ in decoding è necessario riportarla alla dinamica originale

Formula di Dequantizzazione (Long block)

- **SF_mult** = fattore di amplificazione della dinamica dei Fattori Scala
 - se vale 1 equivale ad un raddoppio della dinamica (2^{SF})
 - se vale 0 equivale ad un aumento logaritmico della dinamica ($\sqrt{2}^{SF}$)
- **ScaleFactor** = pesi associati a gruppi di valori frequenziali, raggruppati secondo la scala di Bark (bande critiche) che servono a “colorare” il rumore di quantizzazione
 - $is[i] = xr[i] * 2^{(-SF_multiplier * SF)}$
- **Global_Gain** = passo di quantizzazione con cui vengono quantizzate le linee frequenziali
 - $|is[i]| * 2^{(1/4 * (GG - 210))}$;

Formula di Dequantizzazione (Short block)

- Per i blocchi corti MPEG fornisce una formula leggermente differente da quella prevista per i blocchi lunghi;
- essa permette di definire passi di quantizzazione adattabili ad ognuno dei tre sottoblocchi del granulo (SubBlock Gain - SBG)

Encoder - Quantizzatore non Lineare

Quantizza le linee spettrali con una formula inversa a quella di dequantizzazione affinché:

- il rumore di quantizzazione introdotto resti al di sotto della soglia di udibilità fornita dal modello psicoacustico (Signal to Mask Ratio – SMR). Ciò viene fatto aumentando i fattori scala ed, ove necessario, lo SF_multiplier
- i valori delle 576 linee frequenziali del granulo siano codificabili con le tabelle Huffman fornite dagli standard ISO/IEC 11172-3 e 13818-3; in caso negativo, l'encoder diminuisce l'intera dinamica dello spettro incrementando di 1 il Global_Gain

Differenti approcci

- 1. Manipolazione eseguita prima del blocco di quantizzazione (con modello psicoacustico)**
 - Migliore qualità audio ottenuta (riquantizzazione controllata da modello psicoacustico)
 - Elevato carico computazionale
- 2. Manipolazione eseguita modificando il blocco di quantizzazione (senza modello psicoacustico)**
 - Qualità audio non controllabile dinamicamente; va gestita a priori nella progettazione degli algoritmi
 - Basso carico computazionale

Approccio 1

- Con tale approccio si agisce sullo spettro dequantizzato (no GG, SF, SF_multiplier, SBG).
- Durante le fasi decodifica e ricodifica si evita l'utilizzo del banco di filtri ibrido (polifasico + MDCT)
- Operativamente:
 - Si decodifica l'MP3 fermandosi immediatamente dopo il blocco di dequantizzazione
 - coerentemente con l'input fornito dall'utente (bitrate, tipo e parametri di manipolazione), si aumentano o diminuiscono le componenti frequenziali
 - Si ricalcola l'SMR del nuovo spettro
 - Si ricodifica il segnale generando un bitstream MP3 ISO-compliant

Approccio 2

- Con tale approccio si modifica il comportamento del quantizzatore non lineare agendo, sia sui parametri della formula di quantizzazione che sulle linee frequenziali quantizzate.
- Durante le fasi decodifica e ricodifica si evita l'utilizzo del banco di filtri ibrido (polifasico + MDCT) e del modello psicoacustico
- Operativamente:
 - Si decodifica l'MP3 fermandosi immediatamente dopo la dequantizzazione Huffman
 - coerentemente con l'input fornito dall'utente (bitrate, tipo e parametri di manipolazione), si modificano opportunamente le componenti frequenziali ed i parametri di quantizzazione
 - Si ricodifica il segnale generando un bitstream MP3 ISO-compliant

Modifica Volume

- In entrambi gli approcci lo spettro è uniformemente manipolato in modo da ottenere un bitstream ISO-compliant in cui il loudness del segnale codificato risulti incrementato o decrementato rispetto all'originale

Filtraggio

- In entrambi gli approcci le linee frequenze e/o i fattori ascala ad essi associati sono manipolati in funzione del in funzione del tipo di filtro (LPF, BPF, HPF, BRF) e la frequenza di taglio fornito in input dall'utente.
- Indicativamente, la risoluzione frequenziale dello spettro contenuto in un singolo granulo è pari a:
 - $\text{Res (Hz)} = F_c / 2 * 576$ (F_c = frequenza campionamento)
(La frequenza in Hertz associata ad un generico coefficiente MDCT è pari a 'k*res')

Filtraggio

- Caso semplice: azzerare le frequenze contenute nella banda proibita del filtro fornito dall'utente
- Caso complesso: ampliare o ridurre l'intensità delle frequenze contenute nella banda proibita del filtro fornito dall'utente

Filtraggio ed Alias Polifasico - MDCT

- In fase di analisi l'MDCT ed il banco di filtri polifasico introducono degli alias, eliminati automaticamente in fase di resintesi (TDAC - Time Domain Alias Cancellation)
- La modifica dello spettro deve avvenire in maniera tale che il sistema per l'eliminazione dell'alias di Polifasico - MDCT continui ad operare correttamente

Fase di ricodifica

- Possono presentarsi due problemi:
 - esiste almeno un valore frequenziale che non sia codificabile con le tabelle Huffman
 - Il valore di bitrate fornito in input non è sufficientemente elevato per codificare le linee frequenziali manipolate
- Tali problemi possono essere risolti nei seguenti modi:
 - Aumento del fattore scala corrispondente alla frequenza non codificabile e conseguente diminuzione dei valori frequenziali appartenenti alla banda critica
 - Se ciò non è sufficiente: aumento del Global_Gain e conseguente diminuzione di tutti i 576 valori frequenziali
 - Se disattivata, utilizzo della tecnica del Bit Reservoir

MP3 Bitrate Downgrading

MP3 Bitrate Downgrading

- Tradizionalmente eseguito riportando l'MP3 nel dominio non compresso

MP3 (br1) → [decodifica] → PCM → [ricodifica b2] → MP3 (br2) *con br2 < br1*

- Il passaggio nel dominio PCM può introdurre delle degradazioni nella qualità del segnale, maggiormente percepibili a bassi valori bitrate

MP3 Bitrate Downgrading

E' possibile ottenere l'abbassamento del bitrate, in differenti modi, operando direttamente nel blocco di quantizzazione:

- Downmix di MP3 stereo non compressi
 - bitrate dimezzato a scapito di perdita spazializzazione del suono
- Graduale taglio delle alte frequenze
 - Significativa riduzione del bitrate con conseguente perdita di brillantezza del suono
- Taglio delle bande non vocali (solo segnali vocali)
 - Si perde in termini di qualità e brillantezza ma generalmente non viene intaccata l'intelleggibilità.
- Riquantizzazione del bitstream (GG, SF, SFS, SBG)
 - Significativa riduzione del bitrate con possibili (ma dimensionabili) degradazioni del segnale

Progetti attivi al LIM (MP3)

- Analisi
 - Sincronizzazione automatica di partiture XML e tracce audio MP3
- Manipolazione
 - Bitrate Downgrading
 - Ottimizzazione algoritmi: modifica volume, filtri ed EQ
 - Algoritmi di normalizzazione (peak – RMS)
 - Editor audio Object Oriented (OO) per manipolazione diretta
- Qualità
 - Valutazione oggettiva e soggettiva della qualità audio di MP3 codificati e transcodificati in differenti modi

Prospettive e Proposte di Tesi

- Analisi
 - Beat tracking
 - Pitch Tracking
 -
- Editing - Processing
 - Algoritmi di mixing
 - Time Stretching (WSOLA – Beat Detector)
 - Pitch/Frequency Shifting
 - Effetti stereo effects (panning)
 - Effecti speciali: delay (echi, reverberi), flanger, ecc.
 - Transcodifiche MPEG AAC <-> Dolby AC-3
 -
- Qualità
 - Valutazione dei formati di compressione *lossy*.
 - Sviluppo di tecniche per la valutazione oggettiva della qualità audio percepita (PEAQ - Perceptual Evaluation of Audio Quality)
- Website: http://www.lim.dico.unimi.it/didatt/propostetesi_mp3_aac.html

Prospettive e Proposte di Tesi

- Ricerca e sviluppo di tecniche per l'analisi e l'editing diretto codifiche MPEG Layer 3 ed AAC (Audio Advanced Coding)
- Integrazioni di tali tecniche in applicazioni di editing audio (i.e plug-in SoundForge, ecc.), streaming over IP (i.e Voice over IP, Radio Web, ecc.) e Downloading (Peer2Peer, Music on Demand, ecc.)
- Studio di tecniche e metodologie per la valutazione oggettiva e soggettiva della qualità audio percepita da segnali compressi con perdita d'informazione (principalmente MP3, AAC, WMA, OGG, mp3PRO)
- Proposte di tesi:
 - http://www.lim.dico.unimi.it/didatt/propostetesi_mp3_aac.html
 - <http://www.lim.dico.unimi.it/didatt/dispo.html>