

Audio processing: FORMATI 1

■ MPEG: Motion Picture-Coding Experts Group

E' uno standard per i file di tipo multimediale(video,audio)

Definisce standard di

- Compressione
- Decompressione
- elaborazione
- Codifica

MPEG-1 1992

- 300 kbps per audio stereofonico cioè 1-2 canali
- Usa come tasso di campionamento 33,44.1 e 48 kHz (solo questi 3)
- La compressione genera un bitrate da 32 a 224 kbps, un valore usuale è 192 kbps

Vi sono 3 livelli di compressione (versioni)

- Layer 1 : bitrate > 128 kbps
- Layer 2 : bitrate = 128 kbps
- Layer 3 : bitrate = 64 kbps

MPEG-2 1994

Bitrate a 6 Mbps

Ci sono 6 canali surround:

- Sinistro
- Centrale
- Destro
- Sx-dx
- Surround
- 1 subwoofer per le basse frequenze

MPEG-3 è pensato per le TV-HD, usa il 2

MPEG-4 1999

Standard nato per la codifica.

Audio è composto da diversi oggetti indipendenti

codifica in maniera indipendente situazioni diverse: ad esempio comprime in maniera diversa un personaggio che si muove e il background

MPEG-7 2001

Standard nato per archiviazione

E' uno standard per la ricerca, filtraggio e gestione dati, lo fa usando dei metadati.

Usa XML.

Non si occupa di compressione e decompressione

E' orientato al web in quanto permette la ricerca all'interno di un database di un video/audio

Con MPEG-4 viene spesso chiamato MPEG-4.7 per codifica e descrizione

MPEG-21

Standard nato per interfaccia di scambio.

Standard nato per permettere in maniera migliore lo scambio degli elementi digitali multimediale.

Nasce quindi come standard per la definizione di un framework per sviluppare app multimediali.

MPEG-D , parte da MPEG-21

- Parte 1: MPEG Surround
- Parte 2: Spatial Audio Object Coding
- Parte 3: Unified Speech and Audio Coding

PROPRIETA' DELLE CODIFICHE MPEG

- Retrocompatibilità (sempre)

Nell'implementazione c'è libertà ma ci sono alcuni obblighi.

Libertà nella implementazione:

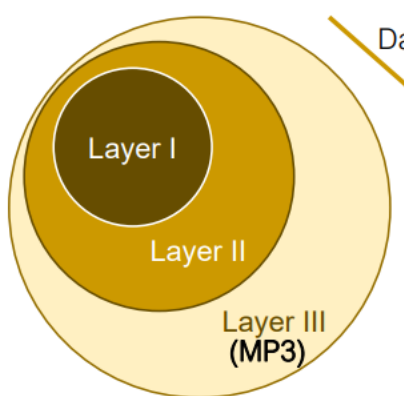
- Obbligatori
 1. Formato dell'audio compresso
 2. Algoritmo di decodifica
- Liberi
 1. Algoritmo di compressione (chiunque usa l'algoritmo di compressione per MPEG deve pagare la società Fraunhofer)

In sostanza per paradosso si potrebbero pagare delle royalties per le parti libere dell'implementazione.

MPEG-1

Abbiamo 3 layer/versioni di compressione

Ogni layer è inglobato nel successivo:



Nei layer successivi abbiamo più compressione ma più lavoro per decoder e encoder.

Il layer 2 ingloba layer 1 quindi c'è una compatibilità.

LAYER 1

Dalla traccia audio prendiamo 384(per finestra) campioni dal dominio del tempo.

Quante sono le finestre? $N_{\text{campioni}} / 384$

Poi passo al dominio delle frequenze con la FFT.

Prendiamo poi 384 campioni di frequenze (spettro discreto con 384 punti)

Applichiamo i filtri polifase ovvero facciamo compressione/codifica per bande.
Ci sono 32 bande di frequenza dove ogni banda ha 12 campioni ($384/32 = 12$)

Quanto è grande una finestra in termini di frequenza?

- $(\text{Tasso_campionamento}/2) / 32$

Le bande sono uniformi, con grandezza relativa (alla frequenza di NYQUIST)
LAYER1 applica una compressione/codifica per bande

A quanto tempo corrisponde lo spettro ?

Esempio: con un tasso di campionamento di 48kHz si ha un periodo di campionamento di 0,000021 secondi, avendo 384 campioni si ottiene: $0,000021 \times 384 = 0,008$ secondi = 8 millisecondi

Se diminuisce il tasso di campionamento ?

La durata aumenta, perché il periodo è inversamente proporzionale alla frequenza.

Esempio: con un tasso di campionamento di 44,1kHz si ha un periodo di campionamento di 0,000026 secondi, avendo 384 campioni si ottiene: $0,000026 \times 384 = 0,0087$ secondi = 8,7 millisecondi

Riusciamo a evitare i pre-echi?

Non sempre, perché il valore ottimale per avere blocchi quasi-stazionari è di 2 millisecondi

In pratica MP1 è sensibile al problema dei pre-echi

Differenza con la scala di Bark

Abbiamo qui 32 bande di larghezza uguale invece di 24 di larghezza variabile come nella scala di Bark.

Le 32 bande non riflettono le bande critiche, paradossalmente avere questo bilanciamento delle frequenze crea uno sbilanciamento nella percezione delle frequenze

Questo è un problema di MP1

Ai 12 bit della banda voglio associare un solo esponente e poi memorizzarne le mantisse.

Uso 6 bit per l'esponente, considerando che un quanto è grande 2dB rappresento al massimo 128 dB

La mantissa avrà un numero variabile di bit, questo permette di comprimere molto quando dentro a una banda la mantissa è (circa) costante. Usiamo 4 bit e quindi 15 classi (anche se ce ne sarebbe 1 non usata) per capire quanto è lunga questa mantissa (15 possibili valori di lunghezza della mantissa). In base alla classe dunque la mantissa del campione è memorizzata con un certo numero di bit

Cos'è una classe di quantizzazione?

Indica come debba essere quantizzato il range dinamico, cioè la grandezza di ogni quanto. In MPEG esistono 15 classi, o configurazioni, con una classe speciale che identifica il silenzio con un unico quanto al valore di intensità 0

Perché esistono classi di quantizzazione diverse?

In base alla classe scelta il numero di bit per banda può variare (da 0 a 15 bit per campione)

Cosa implica un numero di bit variabile?

Non è possibile avere il massimo numero possibile di bit per banda in tutte le bande, poiché il numero di bit totale deve essere uguale al bit rate, che è fissato a 33, 44.1 e 48kHz

Come si sceglie il numero di bit per banda?

Dipende dal rapporto segnale rumore nella banda

DIVISIONE FREQUENZE IN 32

Bande uguali/uniformi con grandezza relativa

Esempio con 48kHz

- Nyquist 24 k
- Banda = $24 \text{ k}/32 = 750$

12 CAMPIONI PER BANDA

- 6 bit: ogni campione è normalizzato rispetto al picco della intera banda (fattore di scala fissato)
- 4 bit: 15 classi (configurazioni) di ri-quantizzazione uniforme
- 0-15 bit per campione

STRUTTURA

Struttura dei bit per la codifica

Header	CRC (ECC)	Allocazione Bit	Fattori di Scala	Campioni per Sottobanda	Dati Aux
variabile	variabile	32 x 4 bit	32 x 6 bit	32 x 12 (=384) da 0 a 15 bit	variabile

I filtri adiacenti sul banco soffrono di sovrapposizione delle bande in modo significativo.

Il segnale viene anche analizzato mediante una FFT (con finestra a 512 punti).

In fase di decoding: si applica il banco di filtri in modalità di sintesi con codifica PCM.

MIGLIORAMENTI A MP1 → MP2

- La risoluzione della FFT: 512 → 1024 punti
- Tassi di campionamento: $32/44.1/48 \rightarrow 16/22.05/24 \text{ kHz}$
- Blocchi più grandi: 384 → 1152 (=3 blocchi MP1)
- Frequenze divise in 3 regioni, riquantizzate con:
 - se frequenza bassa: 16 classi (4 bit)
 - se frequenza media: 8 classi (3 bit)
 - se frequenza alta: 4 classi (2 bit)

Esiste per ogni classe una configurazione 0 per indicare il silenzio per una data regione di frequenza

- Si possono risparmiare bit se regioni di frequenza condividono lo stesso ordine di grandezza

MIGLIORAMENTI MP2 → MP3

- TRASFORMATA DISCRETA DEL COSENO MODIFICATA MDCT: permette partizionamento frequenze simile a quello delle bande critiche
- Quantizzazione non uniforme come mu-law o A-law
- Applicazione fattori di scala anche alle sotto bande
- Codifica di huffman
- Bitrate variabile tra i blocchi

MP3 RIESCE A COMPRIMERE 96 A 1 NEL TELEFONO

MP3 E' UN FORMATO RELATIVO A MPEG-1

ALTRI FORMATI AUDIO

MPEG–AAC : Advanced Audio Coding

- Incluso in MPEG-4
- Supporta fino a 48 canali audio (contro i 2 di MP3)
- Un B.R.=128 kbps in AAC è comparabile a un B.R.=192 kbps in MP3

Dolby AC-3 (Audio Coding)

- Compressione di tipo percettivo

WMA: Windows Media Audio

- Formato proprietario di Microsoft (blackbox)
- Prestazioni migliori di MP3
- **Buone prestazioni sulla musica ma non sulla voce**

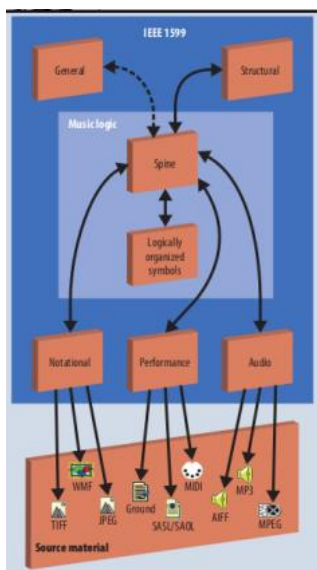
FLAC: Free Lossless Audio Codec

- Compressione lossless (~50%)
- Confrontato con ZIP (~10-20%), è ottimizzato per la compressione della voce
- **LA MIGLIORE PER COMPRESSIONE**

Ffmpeg

Tool che permette di lavorare a basso livello con file di tipo multimediale

IEEE 1599



FORMATO MULTISTRATO BASATO SU XML PER LA DESCRIZIONE MULTIMEDIALE DI OPERE MUSICALI ALL'INTERNO DI UN AMBIENTE SINCRONIZZATO

DEFINITO DA UN SOLO DTD (DOCUMENT TYPE DEFINITION)

GLI STRATI OBBLIGATORI SONO:

- GENERAL
- LOGIC

FORMATI 2

MIDI

MUSICAL INSTRUMENT DIGITAL INTERFACE – nato nel 1983 **ANNI 80**

Rappresenta un linguaggio di comunicazione tra varie apparecchiature, veniva/viene usato ovviamente questo protocollo per l'audio

Nasce per creare uno standard di comunicazione tra apparecchiature di case produttrici diverse che potrebbero usare anche rappresentazioni diverse.

Lo scopo dietro alla sua nascita è stato dunque: **NASCE PER STABILIRE UN'INTERFACCIA COMUNE TRA DISPOSITIVI ETEROGENEI**

SEQUENCER

- Sistema di registrazione ed esecuzione
- Ha memoria programmabile
- Consente editing

Nei sequencer il MIDI è usato come standard per memorizzare i suoni

Il MIDI controlla in modo indiretto il suono, infatti la qualità del timbro non dipende dal MIDI

IL MIDI HA 16 CANALI (definibili)

I CANALI ASTRAGGONO IL CONCETTO DI STRUMENTO

Il MIDI ha un certo numero di canali, da 1 a 16:

- Implementano lo "strumento"
- I messaggi diretti a un canale sono associati tutti allo stesso timbro
- I timbri sono trattabili in modo simultaneo

In generale **LE TRACCE DEFINIBILI NEL MIDI SONO N**

LE TRACCE INDICANO L'ASTRAZIONE CHE RAPPRESENTA LE VOCI DI UN BRANO

LE TRACCE INDICANO L'ASTRAZIONE CHE IMPLEMENTA PARTITURE E MIXING

Le tracce sono da 1 a N

- Implementano partiture e mixing
- Permettono la distinzione logica dei contenuti (esempio: distinguere melodia e accompagnamento)

LE PATCH DEFINIBILI NEL MIDI SONO 128

LE PATCH INDICANO L'ASTRAZIONE CHE IMPLEMENTA IL TIMBRO

Patch da 1 a 128

- Implementano il timbro

In generale **I BANCHI DEFINIBILI NEL MIDI SONO N**

Banchi da 1 a N

- Permettono di incrementare le patch (128 sono pochine)

Vi è una relazione N-a-1 tra:

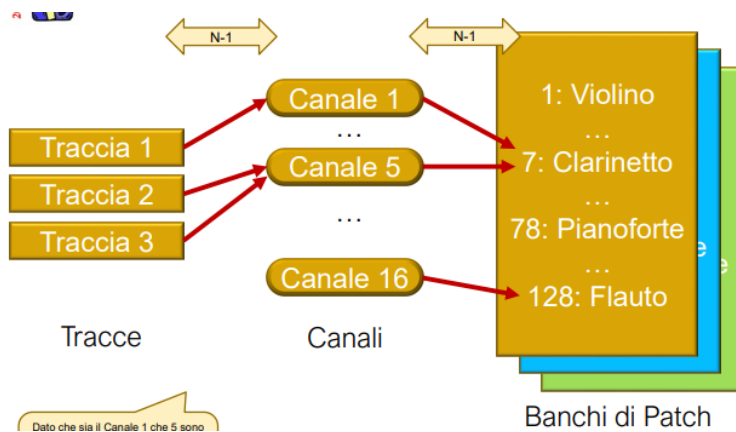
- Tracce-Canali
- Canali-Patch

CANALI-E-TRACCE : 1-N

TRACCE-E-CANALI : 1-N

PATCH-E-CANALI : 1-N

CANALI-E-PATCH : N-1



SCAMBIO DI MESSAGGI

Esempio: suonare una nota

- Ci sono 2 messaggi di inizio e fine dell'esecuzione della nota che sono "Note on" e "Note off"
- Parametro velocity: velocità di completamento della fase di inizio e fine
- Parametro aftertouch: indica variazioni nella pressione (effetto tremolo ad esempio)
- Parametro pitch: indica l'altezza di una nota, sono disponibili 128 valori

I dispositivi MIDI hanno un clock per ordinare i messaggi e usano come unità di misura i tick detti anche PARTI-PER-QUARTO PPQ.

Il valore varia tra 24 e 4096 (i valori assumibili sono sempre multipli o potenze di 2)

I PPQ sono un'unità di misura relativa al Quarto

Quarti per minuto: BEAT-PER-MINUTO BPM. Il valore varia tra 40 e 240

ESERCIZIO: QUANTO DURA 1 TICK?

- **BPM: 120**
- **PPQ: 24**

Un quarto/beat dura

- $60 \text{ secondi} / 120 = 0,5 \text{ secondi}$

Un tick/parte dura

- $0,5 \text{ secondi} / 24 = 0,02 \text{ secondi}$

Il valore assunto da PPQ è detto DIVISION, più è alto più risoluzione temporale c'è

ESERCIZIO(nell'esame): QUANTO DURA 1 TICK?

- **BPM: 240**
- **PPQ: 48**

Un quarto/beat dura

- $60 \text{ secondi} / 240 = 0,25 \text{ secondi}$

Un tick/parte dura

- $0,25 \text{ secondi} / 48 = 0,005 \text{ secondi}$

MESSAGGI MIDI

I messaggi usati sono parole di 10 bit di cui il primo e l'ultimo sono chiamati start e stop.

Allora la parte fondamentale che caratterizza la nota possiamo dire che è rappresentata da 1 byte



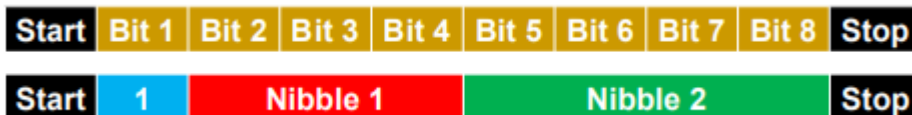
Il bit1 se è 1 indica che si tratta di uno Status byte(identifica tipologie di messaggi), se è 0 indica che si tratta di un Data type(contiene parametri dei messaggi)

Abbiamo 2 tipi di status byte:

- MESSAGGI DI CANALE: ovvero produzione di una nota, cambio timbro per fare degli esempi
- MESSAGGI DI SISTEMA: relativi a sincronizzazione, timing specifiche ecc...

Dividiamo il byte in 3 parti, il primo bit(che sarà 1), il Nibble 1 (3 bit) e il Nibble 2 (4 bit)

Il Nibble se assume valori tra 000 e 110 indica che è un messaggio di canale, se è 111 indica che si tratta di un messaggio di sistema.



A loro volta i messaggi di canale e sistema si suddividono in altre categorie

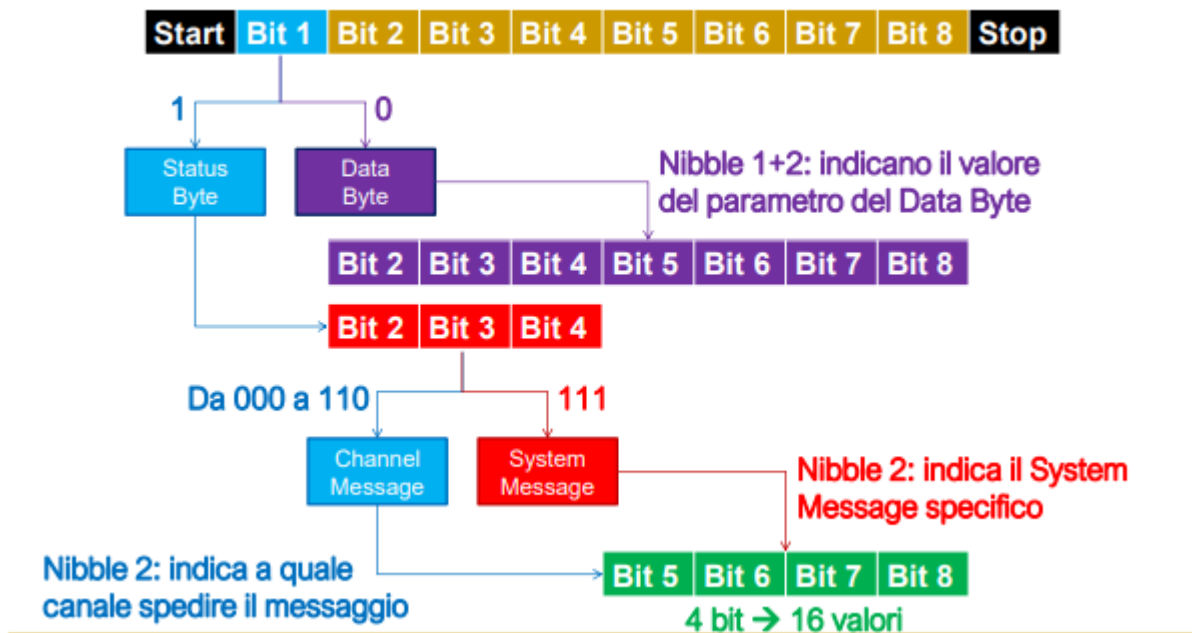
Messaggi di canale:

- MESSAGGI DI VOCE: indica cosa deve suonare uno strumento
- MESSAGGI DI MODALITA': indica il comportamento da assumere rispetto ai messaggi di voce

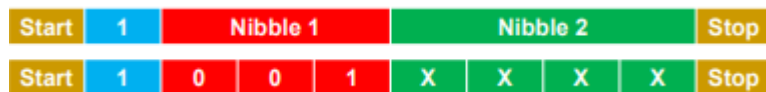
Messaggi di sistema:

- Common message
- Real time message
- Exclusive message

SCHEMA RIASSUNTIVO DEL MESSAGGIO



ESEMPIO: NOTE ON

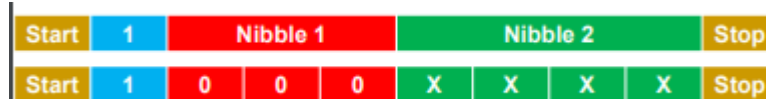


Indica di suonare una nota, il Nibble 2 indica a quale canale mandare il messaggio. Viene mandato in combinazione con 2 Data byte.



Nel data byte avendo a disposizione 7 bit possiamo esprimere 128 valori, un data byte lo usiamo per esprimere il valore di velocity e l'altro per il valore di pitch

ESEMPIO: NOTE OFF



Indica di smettere di suonare una nota, da un punto di vista teorico è come un note on con velocity 0. Anche questo messaggio è accompagnato da una coppia di data byte.



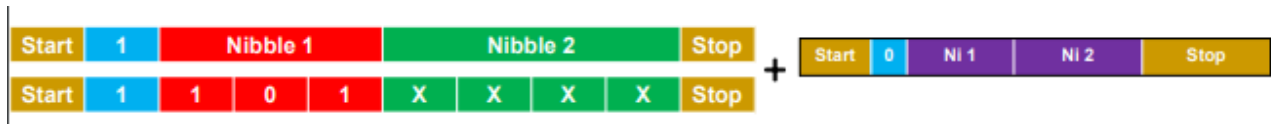
Ottimizzazione di note on e note off

Sostituiamo a Note On e Note Off con:

- Un solo Note On seguito da 2 coppie di data byte la cui seconda coppia ha velocity 0.

Con questa ottimizzazione invece di inviare 6 messaggi ne invieremo 5.

ESEMPIO: CHANNEL PRESSURE (AFTERTOUCH)



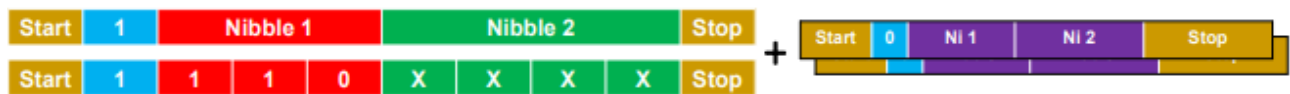
Indica il tremolo.

Il Nibble 2 serve ancora indicare il canale

Il data byte che indica una variazione di pressione(128 valori possibili) solitamente si applica a tutte le pitch/note attive.

Per variare un singolo pitch viene usato il messaggio con Nibble 1 = 010

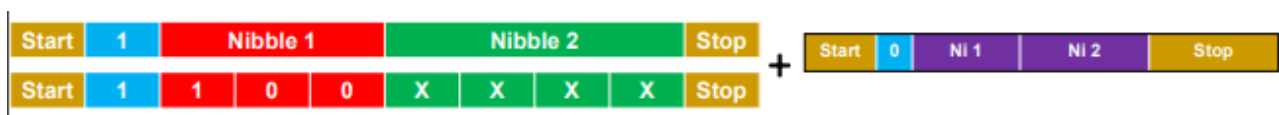
ESEMPIO: PITCH BEND CHANGE



Indica il vibrato.

Usiamo 2 data byte: abbiamo 128*128 possibili valori di variazione in frequenza

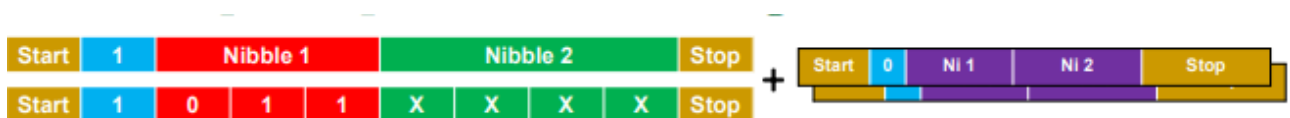
ESEMPIO: PROGRAM CHANGE



Questi messaggi sono usati per il cambio di timbro

Il data byte mette a disposizione 7 bit quindi 128 possibili indici per il cambio del timbro.

ESEMPIO: CONTROL CHANGE



Questi messaggi si riferiscono a vari eventi che producono degli effetti musicali come la pressione sul pedale, il controllo del fiato ecc...

Avendo 2 data byte abbiamo 128*128 valori possibili.

Esempio:

- DataByte#1 : 00000000 indica Cambio Banco
- DataByte#2 : 00000100 indica Banco numero 4

Riassunto dei messaggi di controllo:

- 000 : Note Off
- 001 : Note On
- 010 : Polyphonic Key Pressure
- 011 : Control Change
- 100 : Program Change
- 101 : Channel Pressure (tremolo)
- 110 : Pitch Bend Change (vibrato)

ESERCIZI ESAME

Se i primi 4 bit, esclusi quelli di sincronizzazione, di un messaggio MIDI sono uguali a 1001, allora si avrà un... **Channel Message - Note On**

Se i primi 4 bit, esclusi quelli di sincronizzazione, di un messaggio MIDI sono uguali a 1100, allora si avrà un... **Channel Message - Program Change**

Se i primi 4 bit, esclusi quelli di sincronizzazione, di un messaggio MIDI sono uguali a 1111, allora si avrà un... **System Message**

Se i primi 4 bit, esclusi quelli di sincronizzazione, di un messaggio MIDI sono uguali a 0111, allora si avrà un... **Nessuno dei precedenti** la risposta sarebbe "data byte", essendo il primo bit uguale 0, non è un tipo particolare di messaggio.

I SysEx... **Sono utilizzati dai costruttori per specifiche proprietarie**

MESSAGGI DI SISTEMA (nibble 1 = 111)

COMMON MESSAGE:

- **MIDI Time Code (MTC) Quarter Frame** Per avere un riferimento temporale assoluto
- **Song Position Pointer** Indicizza all'interno di una song (insieme di messaggi)
- **Song Select** Seleziona una song
- **Tune Request** Richiesta di verifica di accordatura (poco usato)

REAL TIME MESSAGE:

- **MIDI Clock** Per avere un riferimento temporale relativo (alla sincronizzazione)
- **Start / Continue / Stop** Gestione attivazione dei sistemi MIDI
- **Active Sensing** Mantiene aperta la connessione fra due device
- **System Reset** Reimposta la configurazione del device

EXCLUSIVE MESSAGE (SYSEX):

- Utilizzati dai costruttori per specificare informazioni relative ai loro prodotti
- Sono gli unici messaggi ad avere 2 Status Byte (apertura e chiusura di un SysEx)
- Se non riconosciuti dai device vengono ignorati

Nota sul nibble 2

Anche se può indicizzare 16 canali, in realtà ne sono stati implementati solo 11(bastavano quelli).

In pratica se dovesse servire spazio per applicare qualche modifica, il protocollo prevede già dei valori non usati.

Evoluzioni di MIDI

General MIDI (GM, o GM1)

- Miglioramento della gestione delle patch (timbri)

General Standard (GS) – Roland

- Aggiunto il messaggio per cambiare banco

Standard XG – Yamaha

- Aggiunta di ulteriori patch (timbri)