



# AUDIO PROCESSING

*UNIVERSITA' DEGLI STUDI DI CATANIA*  
*DIPARTIMENTO DI MATEMATICA E INFORMATICA*  
*LAUREA TRIENNALE IN INFORMATICA*  
*A.A. 2021/22*  
*Prof. Filippo L.M. Milotta*

**ID PROGETTO:** 6

**TITOLO PROGETTO:** Tutto è Relativo

**AUTORE:** Federico Nicotra

## Indice

<b>1. Obiettivi del progetto.....</b>	<b>2</b>
<b>2. Riferimenti bibliografici.....</b>	<b>2</b>
<b>3. Argomenti Teorici Trattati.....</b>	<b>3</b>
3.1 L'armonia, in pillole.....	3
3.1.1 Introduzione.....	3
3.1.2 Le scale.....	4
3.1.3 Gli intervalli.....	4
3.1.4 Gli accordi – Triadi.....	5
3.1.5 Gli accordi – Quadriadi .....	7
3.2 Applicazioni in campo informatico .....	8
3.2.1 Chord Recognition e riconoscimento della tonalità.....	9
3.2.2 Riconoscimento delle canzoni .....	9

# 1. Obiettivi del progetto

Il progetto realizzato si pone come obiettivo quello di dare una introduzione all'armonia dando alcune definizioni importanti, presentare e analizzare dal punto di vista armonico gli accordi a tre e quattro note, classificandoli, e presentare alcune applicazioni in ambito informatico di queste nozioni. Nello specifico si presenterà il primo algoritmo creato per il riconoscimento degli accordi, si citeranno gli sviluppi che ne sono conseguiti e successivamente si presenteranno i principi di funzionamento delle applicazioni di riconoscimento di brani.

# 2. Riferimenti bibliografici

<http://coding-geek.com/how-shazam-works/> (2015)

- Riconoscimento delle canzoni

<https://armoniainline.altervista.org/armonia-lezioni/armonia-lezione-10.html> (2016)

- Descrizione dettagliata delle settime

[Armonia – Walter Piston](#) (1996)

- Approfondimenti per l'introduzione di armonia

Progettazione e Sviluppo di un Metodo di Rilevamento e Riconoscimento di Accordi  
– Luca Garro (2021)

- Algoritmi di riconoscimento degli accordi, storia e dettagli di funzionamento

Analisi della struttura armonica – Emanuele Vassallo (2021)

- Riconoscimento di accordi e tonalità.

### 3. Argomenti Teorici Trattati

#### 3.1 L'armonia, in pillole

##### 3.1.1 Introduzione

Ormai da centinaia di anni la musica occidentale si basa su strutture armoniche ben conosciute e riconoscibili. La teoria musicale ad esse correlate è molto vasta ma ne daremo solo un'introduzione basilare.

Quando si parla di armonia musicale si intende studiare la relazione che esiste tra le note, distinguendole secondo la loro distanza e il modo in cui vengono suonate, e coinvolge la costruzione di accordi e melodie (successioni di note).

Innanzitutto, una **nota** è un segno grafico usato per riferirsi ad un determinato suono, ad una specifica frequenza. Considerando un sistema a temperamento equabile e le possibili alterazioni (simboli che modificano l'altezza della nota, la sua frequenza), individuiamo, in ordine, 12 note:

- |                     |                      |
|---------------------|----------------------|
| 1) DO               | 7) FA# / SOL $\flat$ |
| 2) DO# / RE $\flat$ | 8) SOL               |
| 3) RE               | 9) SOL# / LA $\flat$ |
| 4) RE# / MI $\flat$ | 10) LA               |
| 5) MI               | 11) LA# / SI $\flat$ |
| 6) FA               | 12) SI               |

È altresì importante sapere che nello spettro delle frequenze le note si ripetono in maniera ciclica teoricamente all'infinito, perché la nota successiva al SI è un DO. Tra due qualsiasi di queste note intercorre una distanza che è pari al numero di note che le separano. La distanza tra note adiacenti è sempre uguale ed è chiamata **semitono**. Fisicamente però il semitono è diverso per ogni nota in quanto esso dipende dalla frequenza di partenza per un fattore  $2^{\frac{1}{12}}$ , se consideriamo un sistema a temperamento equabile. Due semitoni sono chiamati un tono. In maniera particolare la distanza di 12 semitoni individua una **ottava** poiché le note considerate avranno frequenza una doppia dell'altra e avranno lo stesso nome. Questa diventa rilevante quando si fa riferimento alle frequenze associate alle note stesse poiché aggiungendo il numero dell'ottava a cui una nota appartiene possiamo univocamente identificarla all'interno dello spettro, ad esempio il DO<sub>3</sub> è la nota DO appartenente alla terza ottava (130,8 Hz), oppure il LA<sub>4</sub> è la nota LA appartenente alla quarta ottava (440,0 Hz). Nella teoria musicale, questa distanza viene chiamata **intervallo** ed è l'elemento base per l'armonia.

### 3.1.2 Le scale

Le scale più conosciute sono le due scale diatoniche, ciascuna di sette suoni, chiamate scala *maggiore* e scala *minore*. La musica tonale dal '700 a questa parte si basa su di esse. La differenza tra le due sta nella successione di toni e semitoni a partire da un punto dato. Per la scala maggiore abbiamo la seguente successione:

$$T - T - sT - T - T - T - sT$$

Per la scala minore abbiamo:

$$T - sT - T - T - sT - T - T$$

I sette suoni della scala sono chiamati **gradi della scala** e vengono chiamati, in ordine: *tonica* (primo grado), *sopratonica* (secondo grado), *mediante* o *modale* (terzo grado), *sottodominante* (quarto grado), *dominante* (quinto grado), *sopraddominante* (sesto grado) e *sensibile* (settimo grado). Si noti che la scala minore può essere ottenuta considerando come tonica il sesto grado della scala maggiore. Il viceversa è possibile se si considera la modale della minore come tonica della maggiore. Esistono dunque 12 tonalità, e scale, maggiori, una per ogni nota esistente e 12 tonalità minori.

### 3.1.3 Gli intervalli

Ne esistono di due tipi e si identificano a seconda di come le note vengono suonate (o rappresentate sullo spartito). L'intervallo è **melodico** se le note sono suonate una dopo l'altra, creando così una melodia. È invece di tipo **armonico** se queste stesse note vengono suonate contemporaneamente, sviluppandosi in verticale sul pentagramma, definendo così un accordo. In figura 1 è riportato un esempio di intervallo melodico e armonico. Gli intervalli subiscono una ulteriore classificazione in base alla distanza che identificano rispetto alla scala diatonica (successione definita di note). Per portare un esempio: la distanza che insiste tra un DO<sub>4</sub> e un MI<sub>4</sub> corrisponde ad un intervallo di terzo grado maggiore (4 semitoni), così come tra un SOL<sub>3</sub> e un SI<sub>3</sub>; la distanza che insiste tra un DO<sub>4</sub> e un MI<sub>b4</sub> corrisponde ad un intervallo di terzo grado minore (3 semitoni). La tabella sottostante raffigura i principali intervalli con il loro nome e la distanza associata. Si intenda che un mezzo tono sia un semitono.



Figura 1: a sinistra è raffigurato un esempio di melodia, con intervalli armonici; a destra un esempio di armonia, con intervalli armonici

Tipo intervallo	Distanza
Prima giusta	0
Seconda minore	1 sT
Seconda maggiore	1 T
Terza minore / Seconda aumentata	1 T e mezzo
Terza maggiore	2 T
Quarta giusta	2 T e mezzo
Quinta diminuita / Quarta aumentata	3 T
Quinta giusta	3 T e mezzo
Sesta minore / Quinta aumentata	4 T
Sesta maggiore / Settima diminuita	4 T e mezzo
Settima dominante / Sesta aumentata	5 T
Settima maggiore	5 T e mezzo
Ottava giusta	6 T

### 3.1.4 Gli accordi – Triadi

Alla luce di quanto detto, un accordo è una sovrapposizione di due o più note suonate insieme. Ogni nota, componente dell'accordo, costituisce una voce. Gli accordi maggiormente usati e su cui ci concentreremo sono accordi formati da 3 e 4 voci, rispettivamente chiamati tricordi o triadi e tetracordi o quadriadi.

Le **triadi** sono molto importanti poiché fungono da base per gli accordi da quattro voci in su. Una triade è dunque formata da due intervalli armonici di terzo grado, cioè formati da 4 o 3 semitoni. La loro combinazione dà vita a quattro tipologie di triadi. Chiameremo **nota fondamentale** dell'accordo quella nota su cui costruiamo l'accordo. Essa sarà il riferimento. Detto questo, le triadi sono:

- La triade **Maggiore (M)** è costituita da un intervallo di terza maggiore seguito da un intervallo di terza minore. Gli accordi costituiti dalla sola triade maggiore sono detti semplicemente “maggiori”. Rispetto alla fondamentale, le due note formano un intervallo di terza maggiore e uno di quinta giusta. La figura 2

raffigura un accordo di DO Maggiore evidenziando gli intervalli relativi (non rispetto alla fondamentale)

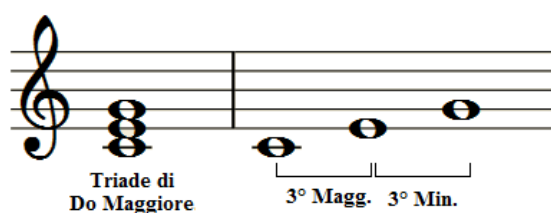


Figura 2: Accordo di DO Maggiore

- La triade **Minore (m)** è costituita da un intervallo di terza minore seguito da un intervallo di terza maggiore. Gli accordi costituiti dalla sola triade minore sono detti semplicemente “minori”. Rispetto alla fondamentale, le due note formano un intervallo di terza minore e uno di quinta giusta. La figura 3 raffigura un accordo di DO Minore evidenziando gli intervalli relativi (non rispetto alla fondamentale)

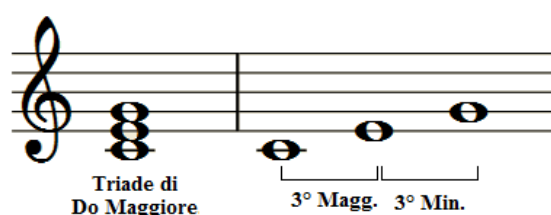


Figura 3: Accordo di DO Maggiore

- La triade **Aumentata** è costituita due intervalli di terza maggiore uno dopo l'altro. Gli accordi costituiti dalla sola triade aumentata sono detti “aumentati”. Rispetto alla fondamentale, le due note formano un intervallo di terza maggiore e uno di quinta aumentata (o eccedente). La figura 4 raffigura un accordo di DO Aumentato evidenziando gli intervalli relativi (non rispetto alla fondamentale)



Figura 4: Accordo di DO Aumentato

- La triade **Diminuita** è costituita due intervalli di terza minore uno dopo l'altro. Gli accordi costituiti dalla sola triade aumentata sono detti “diminuiti”. Rispetto alla fondamentale, le due note formano un intervallo di terza minore e

uno di quinta diminuita. La figura 5 raffigura un accordo di DO Aumentato evidenziando gli intervalli relativi (non rispetto alla fondamentale)

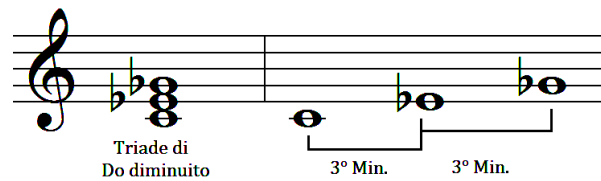


Figura 5: Accordo di DO Diminuito

Alla luce di quanto detto riguardo le scale e le tonalità si riportano di seguito tutti gli accordi di sole triadi possibili per ogni tonalità e per ogni grado della scala (figura 6).

TONALITÀ	I	II	III	IV	V	VI	VII
DO	DO	RE $\flat$ m	MI $\flat$ m	FA	SOL	LA $\flat$ m	SI $\flat$ dim
DO $\sharp$	DO $\sharp$	RE $\sharp$ m	MI $\sharp$ m	FA $\sharp$	SOL $\sharp$	LA $\sharp$ m	SI $\sharp$ dim
RE	RE	MI $\flat$ m	FA $\sharp$ m	SOL	LA	SI $\flat$ m	DO $\sharp$ dim
MI $\flat$	MI $\flat$	FA $\flat$ m	SOL $\flat$ m	LA $\flat$	SI $\flat$	DO $\flat$ m	RE $\flat$ dim
MI	MI	FA $\sharp$ m	SOL $\sharp$ m	LA	SI	DO $\sharp$ m	RE $\sharp$ dim
FA	FA	SOL $\flat$ m	LA $\flat$ m	SI $\flat$	DO	RE $\flat$ m	MI $\flat$ dim
FA $\sharp$	FA $\sharp$	SOL $\sharp$ m	LA $\sharp$ m	SI	DO $\sharp$	RE $\sharp$ m	FA $\sharp$ dim
SOL	SOL	LA $\flat$ m	SI $\flat$ m	DO	RE	MI $\flat$ m	FA $\flat$ dim
LA $\flat$	LA $\flat$	SI $\flat$ m	DO $\flat$ m	RE $\flat$	MI $\flat$	FA $\flat$ m	SOL $\flat$ dim
LA	LA	SI $\sharp$ m	DO $\sharp$ m	RE	MI	FA $\sharp$ m	SOL $\sharp$ dim
SI $\flat$	SI $\flat$	DO $\flat$ m	RE $\flat$ m	MI $\flat$	FA	SOL $\flat$ m	LA $\flat$ dim
SI	SI	DO $\sharp$ m	RE $\sharp$ m	MI	FA $\sharp$	SOL $\sharp$ m	LA $\sharp$ dim

Figura 6: Elenco di triadi presenti per le 12 tonalità maggiori (*m* significa minore, *dim* significa diminuito, maggiore altrimenti)

### 3.1.5 Gli accordi – Quadriadi

Le **quadriadi** sono degli accordi a quattro voci. Per la loro composizione si parte sempre da un accordo di triade che farà da base per il nuovo accordo. Il suo nome dipenderà dall'intervallo che ci sarà tra la nota fondamentale e la nuova nota aggiunta. Principalmente troviamo i seguenti tipi di quadriadi:

- “add2”: si aggiunge una voce che forma un intervallo di seconda maggiore con la fondamentale;
- “add4”: si aggiunge una nota che forma un intervallo di quarta giusta con la fondamentale;
- Sesta: come per le precedenti, si aggiunge la nota che forma un intervallo di sesta maggiore;
- Settima: stesso discorso ma con un intervallo di settima (diminuita, dominante o maggiore)

Le triadi di partenza sono quattro, perciò abbiamo moltissime possibilità di estendere l'accordo con altre note. Per la teoria musicale avremmo appena iniziato, perché ci sono accordi che sono costruiti sugli accordi di settima (come le none, le undicesime e le tredicesime). Tuttavia, quel grado di complessità non è richiesto ai fini del progetto.

Analizziamo nel dettaglio gli accordi di settima, considerati tutti dissonanti. Essi sono divisi in sette tipologie chiamate **specie**. Analizziamole (gli intervalli sono riferiti alla fondamentale):

- *Quarta specie*: formata dagli intervalli di 3° maggiore, 5° giusta e 7° maggiore. Gli intervalli relativi sono: 3° M, 3° m, 3° M. Si forma sul primo o sul quarto grado di una scala maggiore e sul sesto grado di una minore;
- *Prima specie*: formata dagli intervalli di 3° maggiore, 5° giusta e 7° dominante. Questi accordi sono anche chiamati di settima di dominante. Gli intervalli relativi sono: 3° M, 3° m, 3° m. Si forma sul quinto grado di una tonalità maggiore o minore;
- *Sesta specie*: formata dagli intervalli di 3° minore, 5° giusta e 7° maggiore. Gli intervalli relativi sono: 3° m, 3° M, 3° M;
- *Seconda specie*: formata dagli intervalli di 3° minore, 5° giusta e 7° dominante. Gli intervalli relativi sono: 3° m, 3° M, 3° m. Si costruisce o sul secondo o sul terzo o sul sesto grado di una scala maggiore e sul quarto grado in una minore;
- *Settima specie*: formata dagli intervalli di 3° maggiore, 5° eccedente e 7° maggiore. Gli intervalli relativi sono: 3° M, 3° M, 3° m;
- *Terza specie*: formata dagli intervalli di 3° minore, 5° diminuita e 7° dominante. Gli intervalli relativi sono: 3° m, 3° m, 3° M. Si costruisce sul settimo grado della scala maggiore e sul secondo della minore;
- *Quinta specie*: formata dagli intervalli di 3° minore, 5° diminuita e 7° diminuita. Questi sono anche chiamati di settima diminuita. Gli intervalli relativi sono: 3° m, 3° m, 3° m;

### 3.2 Applicazioni in campo informatico

Ci sono stati moltissimi studi che hanno avvicinato l'ambito musicale con quello informatico, dando vita ad algoritmi e applicazioni di uso ormai comune. Basti pensare alle applicazioni di riconoscimento dei brani musicali come SoundHound o Shazam. Esistono anche algoritmi di riconoscimento degli accordi, implementati in applicazioni come Chord AI e Chordify.



### 3.2.1 Chord Recognition e riconoscimento della tonalità

I primi studi per riconoscere gli accordi risalgono al 1999, con l'**algoritmo di Fujishima**. Si applica su una finestra temporale dell'audio che si considera. Si applica dunque una Trasformata Discreta di Fourier (DFT) per ottenerne lo spettro. Questo viene poi "riassunto" nelle 12 note della scala diatonica sommando le intensità delle note, ignorando le ottave. Quindi, sommo le intensità di tutti i DO tra loro, tutti i DO# tra loro, e così via fino al SI. Se volessimo graficare il risultato otterremo un grafico a barre che rappresenta l'intensità totale di ogni singola nota nello spettro. A questo punto, per ultimare il processo d'identificazione, si applicano dei pattern di riconoscimento specifici per gli accordi, chiamati Chord Type Templates, che associano il nome dell'accordo a vettori d'intensità noti. Però questo algoritmo risulta essere sensibile ai rumori di fondo delle registrazioni e ad un problema intrinseco degli strumenti acustici, ovvero che oltre alle note dell'accordo lo strumento produce altre note, poco percepibili, dovute all'acustica e alle imperfezioni dello strumento, che mandano in confusione lo step di identificazione dell'accordo.

Negli anni a venire diversi studi hanno migliorato l'algoritmo di Fujishima, come quello di **Lee e Slaney**, nel 2006, perseguendo la stessa strada dell'originale. Altri hanno totalmente cambiato approccio, anche a fronte di nuove tecnologie, implementando nuovi algoritmi basati su reti neurali, come l'**algoritmo di Humphrey – Bello**, sviluppato nel 2012.

Un passo avanti che sarebbe possibile fare è quello di riconoscere la tonalità del brano, ovvero quale scala diatonica rispetta, non solo capire se maggiore o minore (il modo) ma anche capire quale sia il punto di riferimento, la tonica della scala. Per riconoscere in modo efficiente la tonalità sarebbe l'ideale se si riconoscessero almeno quattro accordi differenti poiché altrimenti ci sarebbe ambiguità. Si considerino solo tre accordi, come RE, LA e FA#m. Essi sono presenti sia nella scala di RE maggiore che in quella di LA maggiore. Alternativamente si evita ambiguità se si riconoscono un accordo diminuito (unico per scala), o due accordi maggiori consecutivi (es. FA e SOL per la tonalità di DO), o due accordi consecutivi di cui il primo minore e il secondo maggiore (es. FA#m e SOL per la tonalità di RE), o un accordo di settima di prima specie.

### 3.2.2 Riconoscimento delle canzoni

Riconoscere una canzone per noi è semplice, basta ascoltarne qualche secondo e immediatamente siamo in grado di individuarla. Tuttavia, per una macchina non è così semplice. Applicazioni come SoundHound e Shazam affrontano il problema in maniera intelligente. Si parte sempre dallo spettro di una registrazione. Si crea una sorta

di impronta dello spettro (figura 7), marcando con un punto le note più forti nel momento in cui vengono mostrate sullo spettrogramma, evidenziandone l'altezza, la frequenza. Questo permette non solo di abbassare da tre a due le dimensioni la struttura dati su cui si sta lavorando, ma permette anche di diminuire drasticamente la quantità di dati da elaborare, oltre che eliminare eventuali disturbi come i rumori di fondo. Ogni canzone presente nei loro database è rappresentata da una serie di queste impronte, una dopo l'altra. Le applicazioni nei cellulari non fanno altro che ricreare questi pattern e inviarli ai server per confrontarli con quelli già esistenti. Solo che qua sorge un altro problema: esistono migliaia di brani, e centinaia di migliaia di impronte da confrontare, per ogni registrazione. Sarebbe impensabile confrontare in maniera diretta l'impronta appena creata con l'intero database sperando di trovare un match in tempi ragionevoli. Invece, la soluzione impiegata fa uso di hash calcolati in base a quali note sono presenti e a come queste sono separate temporalmente dalle altre. Questo meccanismo si rifà molto a come il nostro cervello riconosce una melodia.

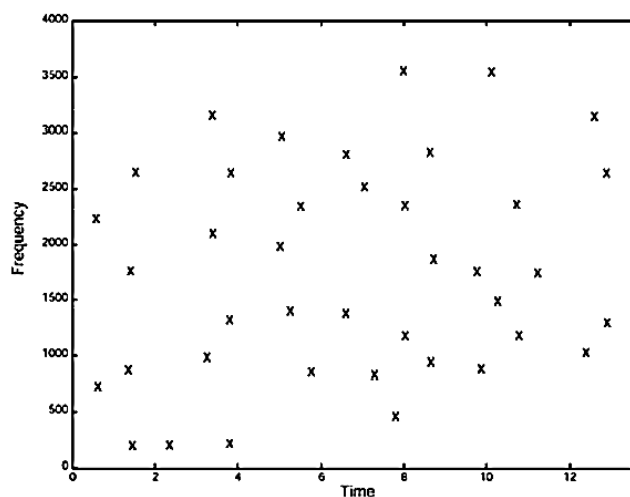


Figura 7: Impronta dello spettrogramma