



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



**DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE**

**CORSO DI LAUREA IN INGEGNERIA DELL'INFORMAZIONE**

**“EFFETTO DELL'ESPRESSIVITÀ SULLA VALUTAZIONE DI MELODIE  
GENERATE COMPUTAZIONALMENTE”**

**Relatore: Prof. Antonio Rodà**

**Laureando: Francesco Colotti**

**Correlatore: Dott. Filippo Carnovalini**

**ANNO ACCADEMICO 2021 – 2022**

**Data di laurea 20/07/2022**



## Sommario

Gli sviluppi recenti nel campo della creatività computazionale hanno portato alla nascita di sistemi avanzati per la creazione di opere artistiche e musicali, con numerose applicazioni apprezzate anche dal grande pubblico. La tesi si restringe all'ambito della creatività computazionale applicata alla musica, in particolare al problema della valutazione delle melodie generate da sistemi, e ha come obiettivo lo studio dell'effetto che una performance espressiva, aggiunta usando degli algoritmi specifici, possa avere sui giudizi di chi ascolta.

Lo studio è stato effettuato impiegando un pubblico di ascoltatori scelto senza discriminazioni basate sulla preparazione musicale, modalità usata da molti ricercatori nel campo. A tale pubblico è stato proposto un questionario dove veniva chiesto di valutare la qualità di alcune melodie, in parte composte da esseri umani e in parte generate, e la plausibilità che ciascuna fosse stata scritta da una persona. Per ogni partecipante metà delle tracce, scelte casualmente con il solo vincolo che fossero per metà composte da esseri umani e per l'altra metà generate, sono state sostituite dalla loro versione espressiva.

Tutto ciò è stato fatto per comprendere se l'espressività possa alterare significativamente i giudizi di un ascoltatore su una melodia generata computazionalmente.

I risultati mostrano che la presenza delle performance espressive non è sempre un fattore significativo nella valutazione delle melodie.



## **Abstract**

The recent developments in the field of computational creativity led to the birth of advanced systems for the creation of artistic and musical works, with numerous applications available to and loved by the general population as well. This thesis focuses on the area of musical computational creativity, especially on the problem of evaluating systems-generated melodies, with the study of the effect of algorithmic expressive performances on listeners' judgements as its objective.

This study has been carried out employing an audience of listeners chosen with no music-expertise-based discrimination, a method used by many researchers in the field. Each listener had to complete a questionnaire where they were asked to rate the quality of a set of melodies, in part composed by humans and in part generated by algorithms, and the plausibility that each of them had been written by a human being. Each participant had half of the tracks replaced by their expressive variants. The replaced tracks were randomly selected, with the sole rule being half of them had to be composed by humans and the other half generated.

All of this has been done to comprehend whether expression can significantly alter one's judgements on a computationally generated melody.

Results show that the presence of expressive performances is not always a significant factor in the evaluation of melodies.



# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Stato dell'arte</b>	<b>5</b>
2.1	Valutazione della musica generata . . . . .	5
2.1.1	Meccanismi di autovalutazione . . . . .	5
2.1.2	Approcci simili al Test di Turing . . . . .	6
2.2	Creazione di performance espressive . . . . .	7
2.2.1	Definizione e rilevanza delle performance espressive . . . . .	7
2.2.2	Tipi di sistemi in uso . . . . .	8
<b>3</b>	<b>Metodo</b>	<b>9</b>
3.1	Ipotesi . . . . .	9
3.2	Strumenti utilizzati . . . . .	10
3.2.1	Director Musices . . . . .	10
3.2.2	Open Science Framework . . . . .	11
3.3	Raccolta dei dati . . . . .	12
3.3.1	Scelta delle melodie: la preselezione . . . . .	12
3.3.2	Aggiunta della performance espressiva con Director Musices . . . . .	13
3.3.3	Raccolta dei dati: il questionario . . . . .	13
<b>4</b>	<b>Analisi e discussione dei risultati</b>	<b>17</b>
4.1	Dati raccolti . . . . .	17
4.2	Analisi dei dati raccolti . . . . .	18
4.3	Discussione dei risultati . . . . .	20
<b>5</b>	<b>Conclusioni</b>	<b>23</b>
5.1	Progetti futuri . . . . .	24
	<b>References</b>	<b>25</b>





# Capitolo 1

## Introduzione

A partire dalla seconda metà del Novecento, lo sviluppo di tecniche algoritmiche sempre più raffinate ha portato diversi esperti a concentrare la propria attenzione su tentativi di ricreare processi creativi tipici dell'uomo negli elaboratori per comprendere più a fondo i meccanismi che regolano la creatività e cercare di stabilire se un computer possa essere creativo [1]. Tale settore di ricerca, che prende il nome di Creatività Computazionale, si è espanso negli anni e ha visto notevoli applicazioni nei campi delle arti. Colton e Wiggins hanno dato una definizione operativa di creatività computazionale come *“la filosofia, scienza e ingegneria di sistemi computazionali che, assumendosi particolari responsabilità, mostrano comportamenti che osservatori imparziali considererebbero creativi”* [2].

Alcuni esempi molto recenti di sistemi di creatività computazionale sono *Dall-e* di OpenAI per le arti visive, il progetto *Aiva* per la generazione di colonne sonore e *Magenta Studio* di Google per la generazione di melodie musicali da quattro battute; anche in ambito accademico la ricerca è molto florida e ha prodotto progetti come *The Painting Fool* di Colton [3], nel campo delle arti visive, e *GenJam* di Biles [4], nel campo musicale.

Nell'ambito della creatività computazionale applicata alla musica, un aspetto storicamente arduo è la valutazione della creatività degli algoritmi, tanto che è stata spesso una fase compiuta in maniera insufficientemente approfondita [5]; la spiegazione di questo fatto si trova principalmente nella difficoltà di definire oggettivamente parametri di confronto fra sistemi creativi, e quindi nella mancanza di uno standard vero e proprio per la valutazione degli stessi. Per tale motivo una tecnica spesso usata è basata sulla *Consensual Assessment Technique (CAT)* di Amabile [6], uno standard per la valutazione della creatività umana che consiste

nel far valutare ad un gruppo di esperti un insieme di prodotti e considerare creativi solo quelli che hanno ricevuto buone valutazioni senza enormi differenze fra i giudici: introducendo fra gli oggetti del confronto anche alcune creazioni di sistemi computazionali, è possibile infatti ottenere un buon criterio di confronto fra creatività umana e computazionale.

Tale metodo può essere usato per raccogliere dei giudizi sulla bontà di un algoritmo; è da notare però che, per una maggiore convenienza economica e logistica, spesso si tende a sostituire gli esperti proponendo piuttosto un questionario ad un pubblico non necessariamente composto da musicisti e preferibilmente ampio, per ottenere una buona idea sull'opinione generale.

Questa tesi si inserisce in tale contesto ponendosi come obiettivo lo studio dell'influenza dell'espressività nella valutazione di melodie generate computazionalmente; in particolare l'interesse è verificare se l'aggiunta di una performance espressiva ad una melodia robotica in formato MIDI possa alterare significativamente la percezione di essa. Provare una relazione tra l'espressività e le valutazioni può essere utile per comprendere più a fondo i fattori che influenzano chi ascolta una melodia.

L'ipotesi che si vuole provare è che se l'espressività influenza i giudizi di un ascoltatore su una melodia, allora l'assenza o la presenza di una performance espressiva sarà un fattore decisivo nella predizione della valutazione di una melodia che influenzerà anche il confronto fra le valutazioni di melodie composte da esseri umani e quelle generate da algoritmi.

Per fare ciò, è stato diffuso un questionario dove ad ogni ascoltatore è stato chiesto di valutare in ordine casuale quattro gruppi di tracce e stimare in seguito per ogni traccia la plausibilità che fosse stata scritta da un essere umano. I primi due gruppi erano rispettivamente composti da tracce composte da esseri umani senza espressività e da tracce generate da algoritmi senza espressività, mentre i restanti due erano formati da tracce raggruppate con lo stesso criterio ma con una performance espressiva aggiunta. Questo tipo di approccio viene spesso associato ad un Test di Turing, ma esiste fra i due una fondamentale differenza, in quanto in esperimenti di questo tipo il partecipante non deve porre domande agli agenti con cui si interfaccia per distinguere uomo e macchina ma deve piuttosto raggiungere la medesima conclusione solamente tramite l'analisi del materiale che essi gli forniscono, senza alcuna interazione [7].

Le performance espressive sono state aggiunte alle melodie con l'ausilio del software *Director Musices* [8], sviluppato e distribuito dalla Scuola di Informatica

e Comunicazione del KTH Royal Institute of Technology di Stoccolma, opportunamente impostato in modo da massimizzare la resa con il genere musicale delle melodie.



# Capitolo 2

## Stato dell'arte

Al fine di giustificare le scelte compiute nell'esperimento, in questo capitolo si cercherà di elencare brevemente gli approcci più comuni per la valutazione dei sistemi generatori di musica e i principali tipi di sistemi esistenti per la creazione di performance espressive.

### 2.1 Valutazione della musica generata

Nonostante lo scopo di questa tesi sia principalmente la stima dell'effetto dell'espressività sui giudizi di un pubblico, si ritiene comunque importante citare alcuni dei metodi più utilizzati per valutare la qualità dei sistemi generatori di musica visto che, per raccogliere i dati di interesse, è stato comunque necessario ottenere valutazioni sulle produzioni.

In generale, nel campo della creatività computazionale, la valutazione della musica generata implica sia verificare che la musica sia di buona qualità sia che l'output sia considerabile creativo. Per questo motivo vengono citati nel seguito sia metodi che aiutano nella valutazione della creatività che metodi che, invece, vengono usati nello studio della qualità del materiale prodotto.

#### 2.1.1 Meccanismi di autovalutazione

Un approccio talvolta utilizzato nella ricerca è far descrivere il funzionamento del sistema all'autore dello stesso, sottolineando perché e fino a che punto esso possa essere considerato creativo. Un primo esempio nella letteratura di tale metodo è il concetto del "*Creative Tripod*" (in italiano "Treppiede Creativo") introdotto da Colton [9], che è un modello in cui, graficamente, le tre gambe corrispondono ad

*abilità, apprezzamento e immaginazione*, caratteristiche ritenute fondamentali dall'ideatore e che devono essere abbastanza approfondite per far sì che il treppiede sia stabile - ossia che il sistema sia considerabile creativo. Altri esempi sono i modelli FACE e IDEA [10], il primo dei quali permette di delineare le capacità creative di un sistema rispetto ad alcune componenti specifiche del processo creativo, mentre il secondo dei quali descrive l'impatto del sistema dalla fase di progettazione al momento in cui è in grado di elaborare materiale mediante processi creativi.

Alcune delle critiche comuni mosse a questi metodi sono legate soprattutto al far valutare un sistema al suo creatore, cosa che potrebbe ledere l'imparzialità dei giudizi; ciononostante, l'utilizzo di questi modelli in fase di sviluppo consente di avere una chiara idea della qualità del sistema in costruzione.

### 2.1.2 Approcci simili al Test di Turing

Benché i metodi citati nel paragrafo precedente siano validi per la valutazione dei sistemi creativi, di cui fanno parte anche i sistemi generatori di musica, esistono altre metodologie, basate sul test di Turing, per raggiungere gli stessi scopi. È bene sottolineare nuovamente che esiste una fondamentale differenza fra questi test, definiti *toy test* da Ariza [7], e il gioco dell'imitazione ideato da Turing, in quanto nei primi il partecipante al test non ha un ruolo di interrogatore ma di critico passivo del materiale fornito; è totalmente assente quindi l'interazione fra macchina e uomo che è alla base di un test di Turing.

Una metodologia importante che rientra in questa categoria è quella proposta da Pearce e Wiggins nel 2001 [11], che consiste, dato un corpus di brani di partenza e supposto che un sistema sia in grado di definire dei parametri che tale corpus rispetta, nel far generare alla macchina altre composizioni che soddisfino i parametri richiesti, proporre alcuni pezzi generati mescolati ad alcuni del corpus originale ad un gruppo di ascoltatori, e chiedere ad essi di stabilire per ogni brano se si tratti di una creazione umana; il sistema di generazione è considerato buono se il suo output non riesce ad essere statisticamente distinto dalle tracce umane. Tale approccio però tende a valutare, più che la capacità di un sistema di essere creativo, l'abilità con cui imita delle composizioni umane [1].

Un altro metodo notevole per la valutazione di sistemi generatori di musica è quello utilizzato ad esempio da Eigenfeldt et al. [12], che consiste nel creare un concerto in cui la valutazione dei brani viene eseguita dal pubblico mediante un questionario; è immediato vedere la similarità che si viene a creare con l'approccio

precedente includendo nella scaletta dell'evento sia pezzi generati sia composizioni di persone in carne e ossa.

Una cosa da sottolineare in entrambi i casi è che introdurre un pubblico di persone non necessariamente esperte implica l'introduzione di differenze dovute alle preferenze dei singoli.

## 2.2 Creazione di performance espressive

### 2.2.1 Definizione e rilevanza delle performance espressive

La definizione di *performance espressiva* deriva direttamente dall'esperienza; è semplice notare infatti che, dato un generico spartito musicale, nessun musicista eseguirà mai il brano in questione esattamente come indicato ma introdurrà, più o meno volontariamente, ad esempio piccole variazioni nella durata delle note, cambiamenti nella forza con cui suona tratti diversi e se fisicamente fattibile - si pensi a strumenti ad arco e alla voce umana - piccole differenze di tonalità rispetto a quella indicata. L'insieme di tutti questi fattori consente ad un musicista di *esprimere* il pezzo musicale, da cui il termine *performance espressiva*. L'esecutore inoltre riesce a trasmettere emozioni diverse eseguendo, sempre più o meno volontariamente, variazioni diverse sul brano. Proprio su quest'ultimo aspetto si è concentrata l'attività di ricerca che ha portato a sistemi come CaRo, di Canazza, De Poli e Rodà [13], che, dato un brano, crea una performance adattata al tipo di intenzione espressiva indicata dall'utente.

Può venire naturale domandarsi a questo punto quale sia l'utilità di programmi che creino artificialmente delle performance espressive; la risposta si ottiene considerando il largo utilizzo, a partire dagli anni ottanta del secolo scorso, dei sequencer prima analogici e poi digitali attraverso i quali possono essere create intere partiture o, ad esempio, aggiunte digitalmente a pezzi già esistenti parti suonate da strumenti virtuali campionati e sintetizzatori. In generale il formato privilegiato con cui questi strumenti lavorano è il MIDI [14], che consente di salvare melodie come insiemi di messaggi di controllo e note con dei precisi valori di intensità, tempi di inizio e fine.

I problemi di tale approccio sono due: i sequencer sono creati per privilegiare un posizionamento perfetto rispetto ai battiti del metronomo delle note nel tempo, note che spesso spesso sono suonate sempre con la stessa intensità, e non interpretano il brano come farebbe un essere umano ma riproducono fedelmente le

melodie passate tramite MIDI. Questo comporta che il brano finale avrà un suono robotico e non naturale, motivo per cui in certi generi musicali può risultare utile la creazione artificiale di una performance espressiva per simulare la presenza di un esecutore umano.

### 2.2.2 Tipi di sistemi in uso

Kirke e Miranda [15] hanno analizzato trenta sistemi per la creazione di performance espressive, catalogandoli in sette gruppi in base al funzionamento degli algoritmi alla loro base:

1. Sistemi che non contengono algoritmi di apprendimento
2. Sistemi che apprendono sfruttando la regressione lineare
3. Sistemi basati su reti neurali
4. Sistemi basati su casi e regole
5. Sistemi basati su modelli grafici statistici
6. Sistemi basati su altri metodi di regressione
7. Sistemi basati sulla computazione evolutiva

Il confine principale può essere tracciato nella pratica fra sistemi che non effettuano alcun apprendimento e creano performance espressive seguendo i parametri dati dall'utente e sistemi che, invece, apprendono le caratteristiche di una musica in un certo contesto e in base ad esse decidono autonomamente o semi-autonomamente le modifiche espressive da attuare in altri contesti. Può essere fatta anche un'altra distinzione più specifica, sempre legata all'uso pratico, fra sistemi automatici che lavorano completamente in autonomia e sistemi semi-automatici che richiedono un intervento umano per riuscire ad elaborare correttamente pezzi diversi.

Fa parte della prima categoria Director Musices (DM), il programma usato per creare le performance espressive in questa tesi (descritto più a fondo nella Sezione 3.2.1), che è stato negli anni un modello particolarmente influente, tanto che in molti dei sistemi analizzati da Kirke e Miranda compaiono delle regole stilisticamente simili a quelle di DM.



# Capitolo 3

## Metodo

In questo capitolo si analizzerà ogni passo dell'esperimento effettuato specificando gli strumenti utilizzati e spiegando l'uso che di essi si è fatto. Si formalizzerà per prima cosa l'ipotesi alla base dell'esperimento e si fornirà solo in seguito una descrizione accurata del lavoro svolto.

Si ricorda che lo scopo dell'esperimento è verificare se la presenza delle performance espressive possa alterare in modo considerevole la valutazione di una melodia generata computazionalmente. Per raccogliere tutte le valutazioni è stato preparato e diffuso un questionario online.

A partire da questo punto ci si riferirà alle melodie generate da algoritmi e a quelle composte da esseri umani rispettivamente, per semplicità, come melodie generate e composte/umane.

### 3.1 Ipotesi

L'esperimento è basato sulla seguente ipotesi: se l'espressività influenza i giudizi di un ascoltatore su una melodia, allora la presenza e l'assenza di una performance espressiva saranno dei fattori significativi nella valutazione di una melodia e i confronti fra melodie generate e composte saranno influenzati da esse.

## 3.2 Strumenti utilizzati

### 3.2.1 Director Musices

Il software utilizzato per aggiungere le performance espressive alle melodie in formato MIDI è stato Director Musices [8], sviluppato e distribuito dalla Scuola di Informatica e Comunicazione del KTH Royal Institute of Technology di Stoccolma. Director Musices è un programma che applica modifiche ad una partitura in formato MIDI in base ad alcune *regole* contestuali, attraverso le quali è possibile manipolare ad esempio l'intensità con cui una nota è suonata e la durata delle singole note all'interno di gruppi, con alterazioni applicate in maniera diversa in contesti diversi. Il programma non effettua alcun tipo di apprendimento automatico ma si basa unicamente sulle impostazioni date dall'utente, motivo per il quale, è catalogabile nella prima delle categorie citate nella Sezione 2.2.2.

L'interfaccia principale del programma è presentata nella Figura 3.1; si può notare in essa che la parte inferiore è dedicata alle regole e la parte alta invece allo spartito su cui si lavora.

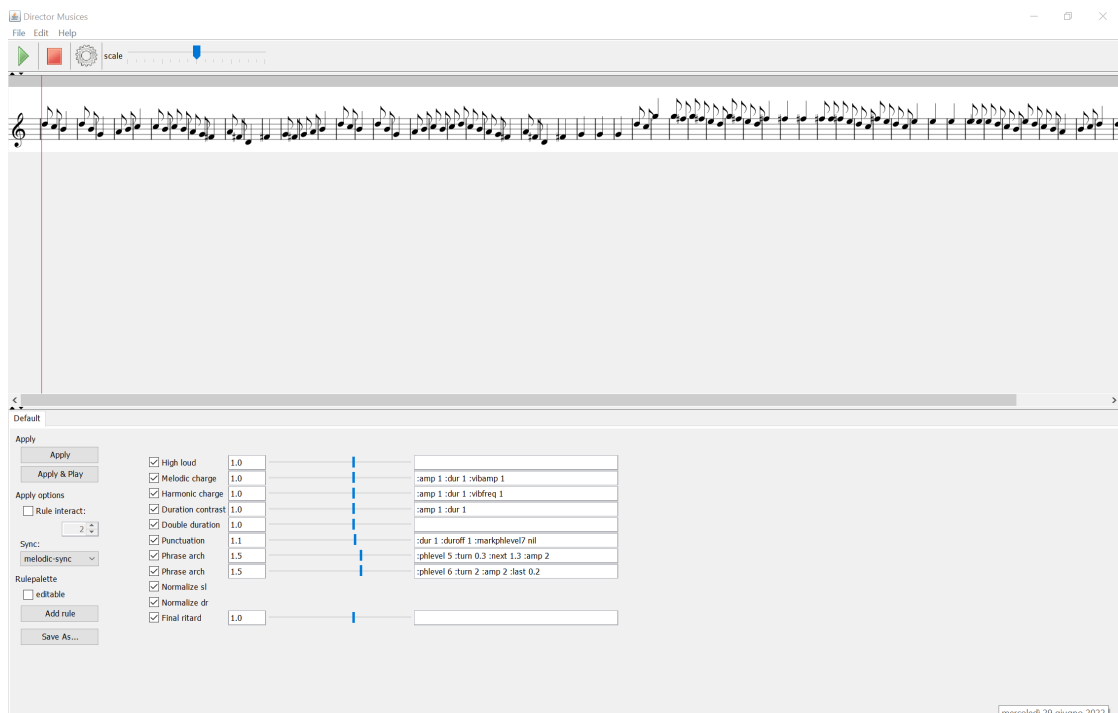


Figura 3.1: La schermata principale di Director Musices

Si è scelto di utilizzare Director Musices perché offre un approccio semplice e intuitivo all'aggiunta di performance espressive, lasciando molto controllo all'u-

tente. Questa libertà, va sottolineato, è da un lato un vantaggio perché consente una maggiore espressione della creatività di chi lo imposta ma da un altro punto di vista uno svantaggio, in quanto l'utente rischia di non applicare le regole in maniera ottimale, creando quindi delle performance espressive scadenti.

Le regole sono in realtà funzioni che, impostate ad un certo valore dall'utente, applicano modifiche contestuali più o meno evidenti ad alcune variabili della performance. Tali variabili sono: il livello del suono, la distanza temporale fra l'inizio di due note successive, la distanza temporale fra la fine di una nota e l'inizio della successiva, ampiezza del vibrato, deviazione in centesimi di semitono dalla tonalità standard dettata dal temperamento equabile (lo standard per l'accordatura degli strumenti nel mondo occidentale) [8].

Due esempi che aiutano a chiarire il concetto sono le regole *High-loud* e *Leap-articulation-dro*. La prima di esse agisce sul volume delle singole note facendo in modo che a note più alte sia associata una maggiore intensità sonora; la seconda invece modifica la distanza fra la fine di una nota e l'inizio della successiva aggiungendo delle micropause nei salti fra note distanti.

### 3.2.2 Open Science Framework

Open Science Framework [16], più brevemente OSF, è uno strumento digitale sviluppato e mantenuto dal *Center for Open Science*, un'organizzazione non a scopo di lucro dedicata a promuovere i valori della trasparenza, dell'integrità e della riproducibilità nella ricerca scientifica. OSF facilita la gestione di progetti di ricerca grazie alla possibilità di organizzare uno spazio condiviso per la gestione di esperimenti o di sottoprogetti in un progetto più grande.

Tra le funzionalità offerte dalla piattaforma, quella utilizzata è stata la registrazione del progetto, che consiste nella creazione di una copia non modificabile in cui sono messe in evidenza data e ora della registrazione. Vengono incluse nella registrazione anche delle copie di tutti i file inclusi nel processo. Una volta registrato, un progetto non può essere né eliminato né modificato ulteriormente ma solo ritirato, azione che ne rimuove i contenuti ma lascia traccia della sua esistenza; questo garantisce un certo grado di trasparenza [17].

Nel caso di questa tesi, è stata compiuta quella che viene più precisamente definita *preregistrazione*, che altro non è che il processo appena descritto ma fatto prima della realizzazione effettiva dell'esperimento/progetto; la preregistrazione contiene, cioè, una descrizione accurata dell'esperimento che verrà effettuato, in

modo tale che ne sia garantita la riproducibilità e la trasparenza. In particolare è stato necessario fornire a priori tutte le informazioni riguardanti l'organizzazione dell'esperimento e la raccolta e l'analisi dei dati, specificando naturalmente le ipotesi di partenza. Preregistrare un esperimento aiuta anche a garantirne la validità in seguito alla sua realizzazione visto che rende più semplice per terze parti controllare che sia stato svolto in maniera coerente con quanto dichiarato in precedenza.

La preregistrazione dell'esperimento svolto in questa tesi è consultabile al riferimento [18].

### 3.3 Raccolta dei dati



Figura 3.2: Fasi del lavoro

Nella Figura 3.2 sono riassunte le fasi del lavoro di raccolta dei dati, che ora verranno spiegate separatamente in dettaglio. Inizialmente si è partiti da un corpus di 24 melodie dello stesso genere musicale, di cui 12 composte e 12 generate. Le melodie composte sono tratte dal Nottingham Musical Database [19], quelle generate invece sono state fornite da Filippo Carnovalini, correlatore di questa tesi. Queste melodie sono state generate per altri esperimenti legati alla ricerca del Dott. Carnovalini [20].

I file audio necessari per lo svolgimento della preselezione e del questionario finale sono stati ottenuti importando le melodie in formato MIDI all'interno di una Digital Audio Workstation (DAW), da cui poi si sono esportate le stesse suonate dalla libreria Noire (che è considerabile come un semplice pianoforte campionato ai fini di questo esperimento) per il campionatore Kontakt.

#### 3.3.1 Scelta delle melodie: la preselezione

La fase di preselezione è stata ideata per scartare metà dei brani iniziali e tenere solo le sei melodie generate e le sei composte più apprezzate da un piccolo gruppo di dieci persone. Questa fase, che è stata necessaria per potersi assicurare di proporre solo il materiale più interessante al pubblico più vasto a cui era rivolto il sondaggio finale, è stata basata su un questionario dal punto di vista funzionale

identico a quello finale. L'unica differenza fra i due è stata la presenza solo nel secondo delle domande riguardanti la plausibilità che le tracce fossero composte da esseri umani e della pagina che richiedeva dati anagrafici generici all'utente. Per questa ragione, per le spiegazioni tecniche sul questionario si rimanda al Paragrafo 3.3.3.

### 3.3.2 Aggiunta della performance espressiva con Director Musices

In questa fase sono state create delle nuove versioni di tutte le tracce non scartate nella preselezione, con l'aggiunta di una performance espressiva. L'insieme delle regole utilizzate e un esempio del risultato ottenuto sono riportati rispettivamente alle Figure 3.3 e 3.4. Tutte le regole sono state impostate cercando di emulare una registrazione, salvata in formato MIDI, di un'esecuzione di una melodia di riferimento su un pianoforte digitale.

Segue ora una breve descrizione delle regole più importanti:

- *Offbeat sl*: impostato a valori negativi rende le note in levare meno intense
- *Leap tone duration* e *Leap articulation dro*: la prima accorcia la prima nota in un salto di tonalità verso l'alto e allunga l'ultima nota in un salto verso il basso, la seconda aggiunge delle micro-pause in tali salti
- *Duration contrast* e *Double duration*: la prima fa sì che note lunghe siano allungate e suonate con più forza e il contrario per quelle corte, la seconda invece, impostata negativamente, aumenta il contrasto delle durate di due note consecutive le cui durate hanno un rapporto 2:1

È possibile leggere una descrizione più accurata della quasi totalità delle regole esistenti in DM, con le indicazioni di tutte le variabili manipolate da ciascuna, all'interno del riferimento [8].

### 3.3.3 Raccolta dei dati: il questionario

La fase della raccolta dei dati, durata circa venti giorni, è stata la parte centrale del lavoro ed è stata svolta con l'ausilio di un questionario online. Il sito del questionario è stato prodotto adattando alle problematiche di questo esperimento un sito già esistente utilizzato per un fine simile.

L'interfaccia del questionario è rappresentata alla Figura 3.5. Il questionario contiene in totale quattro pagine di valutazione, come quella in figura, una pagina di introduzione e una, l'ultima, per la raccolta di alcuni dati anagrafici.

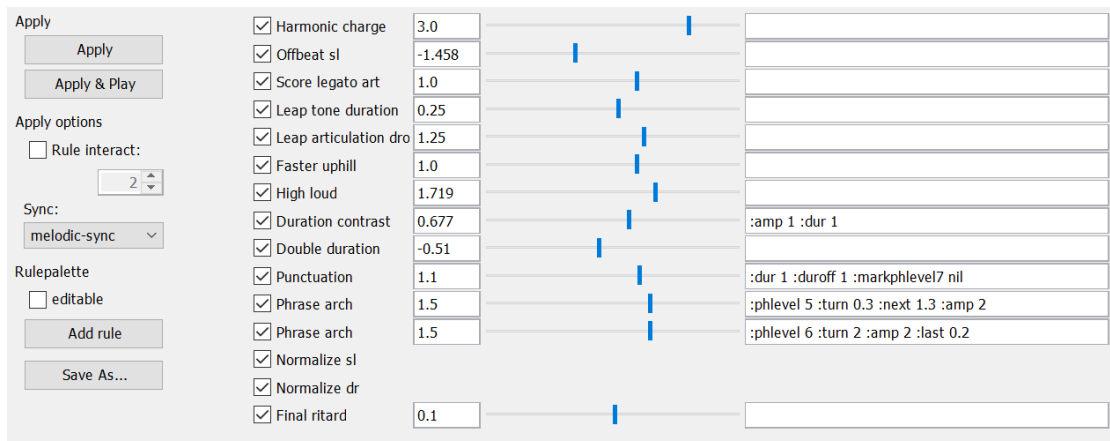


Figura 3.3: Le regole utilizzate

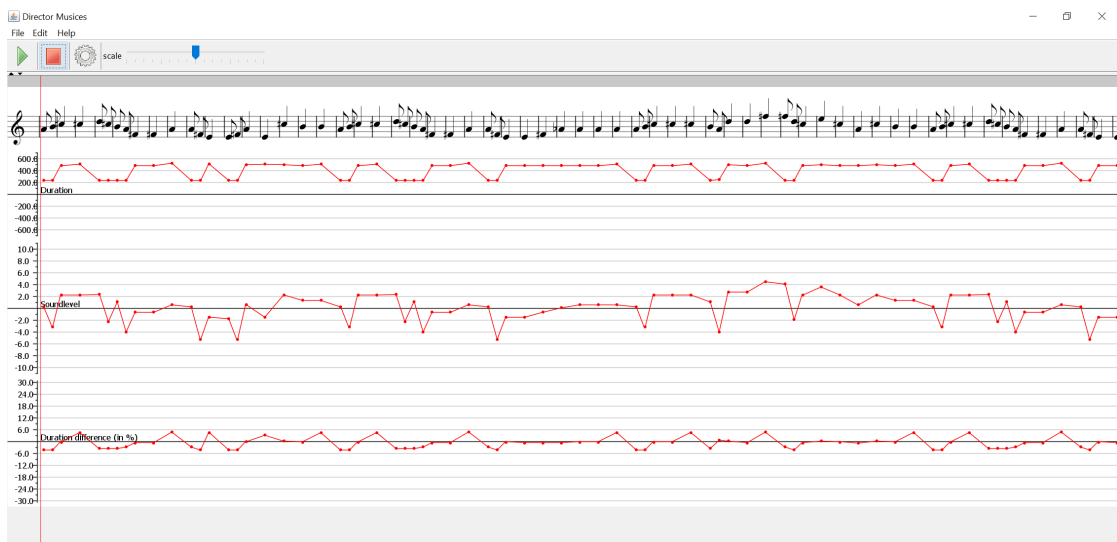


Figura 3.4: Il risultato prodotto dalle regole nella figura 3.3

Al caricamento del sito avvengono due cose: viene inizializzata una variabile con data e ora e vengono creati due array contenenti sei stringhe ciascuno, che corrispondono ai nomi delle tracce audio ( $g_x$  per la  $x$ -esima generata e  $h_x$  per la  $x$ -esima composta). In seguito vengono mescolati gli elementi dei due array, facendo in modo che in ciascuno ci siano tre elementi corrispondenti alle melodie generate e tre a quelle composte; il codice della funzione utilizzata è consultabile nel frammento di codice 3.1. Questo avviene perché ogni partecipante deve ascoltare e valutare tutte le tracce, ma metà di esse devono essere presentate solo nella versione espressiva. In una determinata pagina sono presenti solo tracce con o senza performance. L'ordine in cui le due modalità (con e senza performance espressiva) sono presentate al partecipante è casuale, in modo da ridurre al minimo

## Sondaggio sulla piacevolezza di melodie. Pagina 2 di 6

## Valutazione di melodie

In questa pagina ci sono 6 melodie. Ascoltate tutte e assegna a ciascuna un voto da 1 a 5 (1 = peggiore, 5 = migliore) per indicare quanto la trovi piacevole/bella.

Nel farlo cerca di dare voti relativamente alle melodie su questa pagina e non rispetto alle tue normali preferenze musicali (cioè la melodia che trovi migliore in questa pagina dovrebbe avere voto 5 e il peggiore 1, anche se magari nemmeno la migliore ti piace).

Figura 3.5: L'interfaccia del questionario

ogni possibile bias causato dall'ordine di esposizione, ed è deciso da un "lancio di una moneta" simulato chiamando la funzione *random()* dell'oggetto Math di Javascript.

Si sottolinea che a ciascun utente vengono proposte sempre e comunque in ogni pagina tre melodie composte e tre generate mescolate fra loro, che saranno o in una modalità o nell'altra. A ciascuna modalità corrispondono due coppie di pagine, nelle quali la prima pagina è sempre la valutazione della piacevolezza delle melodie, rispetto a quelle presenti sulla stessa pagina, e la seconda il giudizio sulla plausibilità che ogni melodia sia stata scritta da un essere umano. Entrambe le valutazioni vengono date con una scala Likert a 5 livelli, da molto poco piacevole/molto poco verosimile a molto piacevole/molto verosimile.

```

1 //mixArrays: mescola due array facendo in modo che ci siano
2 //3 elementi dell'uno e 3 dell'altro.
3 //Si dà per scontato i due array abbiano lunghezza 6
4 function mixArrays(a,b){
5   if(a.length != b.length){
6     console.log("mixArrays: the two arrays need to be of the same
7       dimension");
8     return;
9   }
10  let ind = [0,1,2,3,4,5];
11  //shuffleArray è una funzione che mescola casualmente gli elementi

```

```
di un array
11  shuffleArray(ind);
12  for(let i = 0; i < 3; i++){
13    let temp = a[ ind[i] ];
14    a[ ind[i] ] = b[ ind[i] ];
15    b[ ind[i] ] = temp;
16  }
17 }
```

Codice 3.1: La funzione mixArrays

L'ultima pagina del questionario è uguale per ogni utente e in essa viene richiesto di inserire età e genere, lasciando in ogni caso la possibilità di non rispondere, e di indicare il livello di esperienza musicale tramite una singola domanda categorica a sei livelli [21]. In mezzo a queste domande è inoltre presente anche una domanda aggiuntiva a risposta multipla che viene utilizzata come controllo del livello dell'attenzione.

Quando viene cliccato il pulsante per completare il questionario, una funzione crea una tabella contenente tutti i risultati e i dati generici inseriti nell'ultima pagina, oltre che il tempo impiegato, calcolato sottraendo la data e l'ora dell'inizio da quelle della fine. La tabella dei risultati viene infine convertita in una stringa JSON dal metodo `JSON.stringify()` e inviata ad un server per essere memorizzata.



# Capitolo 4

## Analisi e discussione dei risultati

Avendo spiegato approfonditamente il funzionamento del questionario nel capitolo 3, segue ora una parte di analisi dei dati raccolti con esso.

Tutte le operazioni di analisi sono state eseguite con RStudio<sup>1</sup>, un IDE per R, uno dei linguaggi più comunemente utilizzati nel campo dell’analisi statistica. In particolare, per quanto riguarda questo progetto, si è usato RStudio per importare la tabella contenente tutti i dati raccolti dal questionario ed effettuare in modo efficiente le analisi descritte in seguito, nonché per visualizzare tutto ciò con dei grafici.

Nel seguito, ci si riferirà ai voti dati nelle domande di valutazione delle tracce come “valutazioni” e quelli assegnati nelle domande sulla plausibilità che le melodie fossero composte come “plausibilità”.

### 4.1 Dati raccolti

In totale hanno compilato il questionario 37 persone. Di queste, 7 sono state scartate o per aver completato il questionario in meno di un tempo minimo, calcolato supponendo che ogni partecipante ascoltasse almeno una volta ogni traccia, o per aver risposto in maniera errata alla domanda di controllo dell’attenzione. Ciononostante, è stato raggiunto l’obiettivo minimo di 30 questionari validi che ci si era posti nella preregistrazione dell’esperimento [18].

Tutti i dati raccolti sono stati riuniti in una tabella in cui a ogni riga corrisponde un partecipante. Alcuni dati, età e genere, non sono stati utilizzati; per

---

<sup>1</sup><https://www.rstudio.com/products/rstudio/>

completezza, tuttavia, scegliamo di riportare anch'essi: dei 30 partecipanti i cui questionari sono risultati validi, 18 sono uomini, 11 donne e 1 ha preferito non specificare; l'età media è risultata essere di 30.1 anni, con una deviazione standard di 14.8.

Per quanto riguarda invece il livello di conoscenze musicali, 20 partecipanti hanno affermato di essere non musicisti e 10 musicisti; di questi ultimi, 1 si è descritto come “musicista professionista”.

## 4.2 Analisi dei dati raccolti

L'analisi è stata eseguita considerando come variabili indipendenti *author*, l'autore della melodia e *mod*, la modalità. Sono entrambe variabili a due livelli, delle quali la prima può assumere solo i valori *human* e *algo*, indicanti rispettivamente una melodia composta e una generata, e la seconda *deadpan* e *expressive*, che segnalano rispettivamente l'assenza e la presenza della performance espressiva.

Le variabili misurate sono state i voti assegnati, raccolti, come specificato in precedenza, con una scala Likert a 5 livelli. Durante questa parte dell'analisi non si sono mai utilizzati contemporaneamente i voti relativi alle pagine di valutazione e quelli relativi alle pagine di plausibilità.

**Analisi della devianza su modello cumulativo** Per la prima parte dello studio dei dati, si è cercato di adattare un *Cumulative Link Model* (CLM) ai dati ottenuti per predire le valutazioni in scala Likert basandosi sulle due variabili indipendenti e la loro interazione; i giudizi in scala Likert sono dunque stati trattati come fattori ordinali - approccio giustificato dal fatto che tale scala possiede per sua natura delle categorie di valutazioni ben ordinate. In seguito si è eseguita un'analisi della devianza (ANOVA) su tale modello, per valutare la significatività di ciascun predittore.

Questo procedimento è stato eseguito due volte, una per le valutazioni, una per le plausibilità. I risultati per entrambi i casi sono riportati nel seguito sia numericamente, rispettivamente nelle Tabelle 4.1 e 4.2.

Si noti che nelle Tabelle 4.1 e 4.2 *author:mod* rappresenta l'interazione fra le due variabili.

Nella Figura 4.1 è possibile osservare le probabilità per ciascun autore di ottenere i singoli voti nel modello; la Figura 4.1a rappresenta tali dati nel caso delle

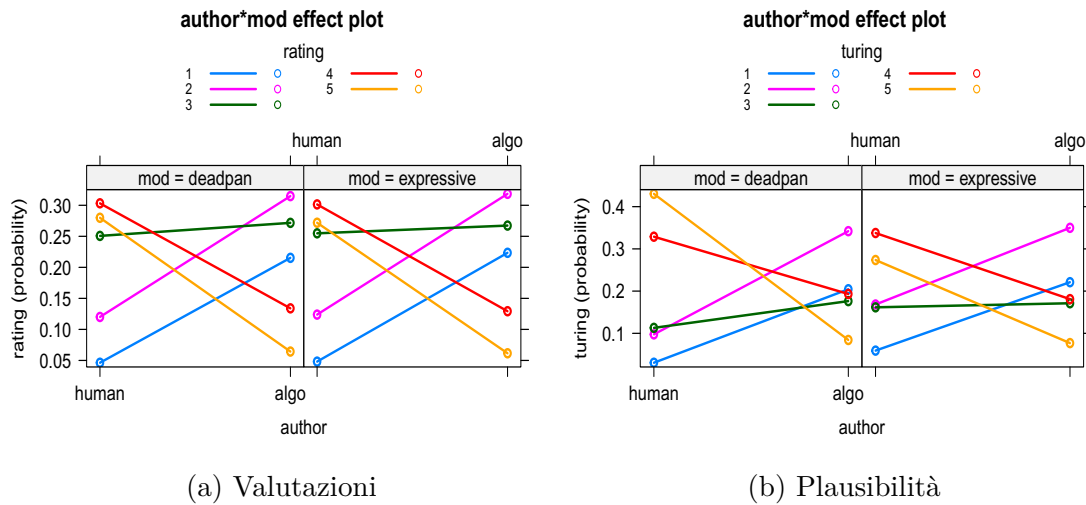


Figura 4.1: I grafici degli effetti dei due modelli, divisi per modalità

valutazioni, mentre la Figura 4.1b nel caso delle plausibilità. In entrambi i casi il confronto è fatto fra i voti senza e con l’espressività.

L’analisi evidenzia una significatività dell’interazione fra autore e modalità solo nel caso delle valutazioni, non delle plausibilità; autore e modalità invece sono in entrambi i casi significativi. Inoltre, pur non essendo riportato nelle tabelle, si segnala che non c’è stata alcuna differenza significativa fra i voti dei musicisti e quelli dei non musicisti.

Oltre a quanto riportato sopra, sono state effettuate delle analisi esplorative non descritte in fase di preregistrazione di cui si descrivono ora i risultati.

**T-test per le medie delle distribuzioni** Per studiare l’esistenza di differenze significative fra le distribuzioni, si è ricorsi al t-test, assumendo che un p-value minore di 0.05 implicasse l’esistenza di differenze significative. I risultati salienti vengono riportati di seguito.

La Tabella 4.3 mostra che ci sono differenze significative nelle distribuzioni dei voti assegnati ai due autori (sia nella parte di valutazione sia nella parte delle

	Df	Chisq	Pr(>Chisq)	
author	1	203.790	<2e-16	***
mod	1	93.807	<2e-16	***
author:mod	1	22.501	2.1e-06	***

Tabella 4.1: Analisi della devianza: valutazioni. “\*\*\*” indica significatività con  $p < 0.001$

	Df	Chisq	Pr(>Chisq)	
author	1	199.6924	<2e-16	***
mod	1	67.7772	<2e-16	***
author:mod	1	0.8138	0.367	

Tabella 4.2: Analisi della devianza: plausibilità. “\*\*\*” indica significatività con  $p < 0.001$

	human	algo
media valutazioni	3.650000	2.483333
media plausibilità	3.783333	2.561111

Tabella 4.3: Medie (senza considerare la modalità)

plausibilità): in generale in entrambe le parti sono stati assegnati voti più alti alle tracce composte che a quelle generate.

Non si sono invece riscontrate, nel corso dell’analisi, differenze significative fra le due modalità né per le valutazioni né per le plausibilità. Anche considerando solo le melodie composte o solo quelle generate il quadro non cambia in generale; l’unica eccezione si nota considerando solo le plausibilità delle melodie umane, dove si osservano (e si ha un p-value di 0.02776) i risultati della tabella 4.4.

	deadpan	expressive
media	3.988889	3.577778

Tabella 4.4: Medie delle plausibilità, rispetto alla modalità, per le tracce composte

La tabella 4.4 mette in evidenza che, per quanto riguarda le distribuzioni delle plausibilità delle melodie composte, le tracce contenenti una performance espressiva sono state ritenute meno plausibilmente umane, nonostante lo fossero. Quanto detto è osservabile anche nella figura 4.1b.

### 4.3 Discussione dei risultati

I risultati ottenuti sono in parte conformi a quanto ci si attendeva, per quanto riguarda la forte relazione fra autore e valutazioni, e in parte molto distanti dalle ipotesi, per quanto attiene invece all’influenza delle performance espressive. Per quanto riguarda l’analisi del CLM, nella parte delle valutazioni è risultata significativa l’interazione fra autore e modalità, come da ipotesi. Nella parte della plausibilità invece non è risultata significativa l’interazione fra le medesime va-

riabili. Osservando il grafico nella Figura 4.1a, si comprende che l'interazione è significativa perché le melodie sono meno apprezzate nella loro versione espressiva, cosa abbastanza inaspettata.

Anche studiando i risultati delle analisi esplorative si nota che la modalità non è particolarmente significativa, e quando lo è, nel caso delle melodie composte, implica un peggioramento nelle valutazioni e nelle plausibilità.

Si può cercare di spiegare quanto osservato in diversi modi. Per quanto riguarda il peggioramento della media delle plausibilità per le tracce umane, una possibilità è che gli ascoltatori, inconsciamente, non si aspettassero l'espressività all'interno di esse e questo li ha portati a considerarle meno plausibilmente umane. Per quanto riguarda invece l'assenza di differenze con l'introduzione delle performance espressive, si potrebbe identificarne una motivazione in una scelta non ottimale delle regole applicate in Director Musices - plausibile, considerando che esse sono state impostate in modo da non rendere troppo evidente l'aggiunta di una performance espressiva artificiale - ma non è nemmeno da escludere che gli ascoltatori percepissero una differenza di qualità troppo alta fra le tracce composte e quelle generate, facendo sì che l'effetto della performance espressiva venisse oscurato dall'effetto principale della differenza di autore.



# Capitolo 5

## Conclusioni

Questa tesi è iniziata con una contestualizzazione del lavoro nel campo della creatività computazionale, spiegando di quali argomenti tratta quest'ultima e citando alcune sue applicazioni in ambito commerciale e accademico. Si è in seguito giustificato l'approccio scelto nella fase della raccolta dei dati presentando i metodi più comuni utilizzati per la valutazione della creatività e delle melodie prodotte da sistemi generatori di musica. Si è poi voluto elencare le categorie a cui appartengono i software che creano le performance espressive, in modo da motivare meglio la scelta dell'uso di Director Musices.

Dopo questa prima parte dedicata all'introduzione e allo stato dell'arte, l'attenzione si è concentrata sulla fase della raccolta dei dati e si sono brevemente descritti l'esperimento effettuato, la preregistrazione, il questionario online e le impostazioni di Director Musices utilizzate. Si sono infine analizzati i dati raccolti e si è giunti alla conclusione che l'ipotesi iniziale, che affermava che l'aggiunta di espressività potesse portare a variazioni significative sulle valutazioni di musica generata, non è stata pienamente confermata. Le analisi esplorative hanno mostrato che, al contrario di quanto ci si aspettava, l'espressività ha implicato un peggioramento dei voti sulla plausibilità sulle musiche composte da esseri umani.

Quanto osservato è stato attribuito o al fatto che i partecipanti non si aspettavano, inconsciamente, l'espressività nelle melodie composte e ciò li ha portati a giudicare le tracce come più plausibilmente generate da un algoritmo, o alla possibilità che sia stata effettuata una scelta non ottimale di regole su Director Musices. Un'altra ipotesi formulata è quella che vi fosse una differenza di qualità percepita troppo alta fra le melodie composte e quelle generate che non è stato possibile colmare con una performance espressiva.

## 5.1 Progetti futuri

Alla luce dei risultati ottenuti, potrebbe essere interessante svolgere nuovamente l'esperimento rendendo più evidenti le performance espressive.

Si potrebbe anche, in un'altra eventuale ripetizione dell'esperimento, scegliere tracce generate da un algoritmo differente, non escludendo che i risultati siano stati condizionati da una differenza di qualità eccessiva fra tracce composte e generate, oppure confrontare due algoritmi ritenuti di qualità simile. Questo può essere sensato perché i risultati ottenuti, sebbene non soddisfacenti rispetto all'ipotesi formulata inizialmente, mostrano che la presenza degli effetti ipotizzati possa essere significativa in condizioni più stringenti di quelle testate in questa tesi.



# References

- [1] F. Carnovalini e A. Rodà. “Computational Creativity and Music Generation Systems: An Introduction to the State of the Art.” In: *Front. Artif. Intell.* 3:14. (2020). DOI: 10.3389/frai.2020.00014.
- [2] S. Colton e Geraint Wiggins. “Computational creativity: The final frontier?” In: *Frontiers in Artificial Intelligence and Applications* 242 (gen. 2012), pp. 21–26. DOI: 10.3233/978-1-61499-098-7-21.
- [3] Simon Colton, Michel Valstar e Maja Pantic. “Emotionally aware automated portrait painting”. In: set. 2008, pp. 304–311. DOI: 10.1145/1413634.1413690.
- [4] John Biles. “GenJam: A Genetic Algorithm for Generating Jazz Solos”. In: lug. 1994.
- [5] Anna Jordanous. “Evaluating Evaluation: Assessing Progress and Practices in Computational Creativity Research”. In: *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems*. A cura di Tony Veale e F. Amílcar Cardoso. Cham: Springer International Publishing, 2019, pp. 211–236. ISBN: 978-3-319-43610-4. DOI: 10.1007/978-3-319-43610-4\_10. URL: [https://doi.org/10.1007/978-3-319-43610-4\\_10](https://doi.org/10.1007/978-3-319-43610-4_10).
- [6] Teresa M. Amabile. “A Consensual Technique for Creativity Assessment”. In: *The Social Psychology of Creativity*. New York, NY: Springer New York, 1983, pp. 37–63. ISBN: 978-1-4612-5533-8. DOI: 10.1007/978-1-4612-5533-8\_3. URL: [https://doi.org/10.1007/978-1-4612-5533-8\\_3](https://doi.org/10.1007/978-1-4612-5533-8_3).
- [7] Christopher Ariza. “The Interrogator as Critic: The Turing Test and the Evaluation of Generative Music Systems”. In: *Computer Music Journal* 33.2 (2009), pp. 48–70.
- [8] Roberto Bresin, Anders Friberg e Johan Sundberg. “Director musices: The KTH performance rules system”. In: *Proceedings of SIGMUS* (gen. 2002).

- [9] Simon Colton. “Creativity Versus the Perception of Creativity in Computational Systems”. In: *AAAI Spring Symposium: Creative Intelligent Systems*. 2008.
- [10] Simon Colton, John Charnley e Alison Pease. “Computational creativity theory: The FACE and IDEA descriptive models”. In: *Proceedings of the 2nd International Conference on Computational Creativity, ICCO 2011* (mag. 2012).
- [11] Marcus Pearce, Geraint Wiggins et al. “Towards a framework for the evaluation of machine compositions”. In: *Proceedings of the AISB’01 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*. Citeseer. 2001, pp. 22–32.
- [12] A. Eigenfeldt, A. Burnett e P. Pasquier. “Evaluating musical metacreation in a live performance context”. In: *Proceedings of the 3rd International Conference on Computational Creativity, ICCO 2012*. 2012, pp. 140–144.
- [13] Giovanni De Poli e Antonio Roda. “CaRo 2.0: An Interactive System for Expressive Music Rendering”. In: *Advances in Human-Computer Interaction 2015* (feb. 2015), pp. 1–13. DOI: 10.1155/2015/850474.
- [14] *Official MIDI Specifications*. URL: <https://midi.org/specification> (visitato il 28/06/2022).
- [15] Alexis Kirke e Eduardo Miranda. “A Survey of Computer Systems for Expressive Music Performance”. In: *ACM Comput. Surv.* 42 (dic. 2009). DOI: 10.1145/1592451.1592454.
- [16] *OSF - Open Science Framework*. URL: <https://osf.io> (visitato il 29/06/2022).
- [17] E.D. Foster e A. Deardorff. “Open Science Framework (OSF)”. In: *J Med Libr Assoc.* 105.2 (apr. 2017), pp. 203–206. DOI: 10.5195/jmla.2017.88.
- [18] Francesco Colotti e Filippo Carnovalini. *The effect of expression on the evaluation of computationally generated melodies*. Giu. 2022. DOI: 10.17605/OSF.IO/HVVG6. URL: <https://osf.io/hvvg6>.
- [19] *Nottingham Music Database*. Il sito contiene, oltre a una versione in formato abc del dataset, un collegamento al sito web originale, non più disponibile, appartenente alla Scuola di Informatica dell’Università di Nottingham. URL: <http://abc.sourceforge.net/NMD/> (visitato il 30/06/2022).

- [20] Filippo Carnovalini et al. “Studying Structural Regularities through Abstraction Trees”. In: *Proc. of the 15th International Symposium on CMMR* (2021), pp. 165–174.
- [21] J. Diana Zhang e Emery Schubert. “A Single Item Measure for Identifying Musician and Nonmusician Categories Based on Measures of Musical Sophistication”. In: *Music Perception* 36.5 (giu. 2019), pp. 457–467. ISSN: 0730-7829. DOI: 10.1525/mp.2019.36.5.457. eprint: [https://online.ucpress.edu/mp/article-pdf/36/5/457/191133/mp\\\_2019\\\_36\\\_5\\\_457.pdf](https://online.ucpress.edu/mp/article-pdf/36/5/457/191133/mp\_2019\_36\_5\_457.pdf). URL: <https://doi.org/10.1525/mp.2019.36.5.457>.

