UNIVERSITÀ DI PADOVA **FACOLTÀ DI INGEGNERIA**

TESI DI LAUREA

EXPRESSIVE INFORMATION PROCESSING: UN'ANALISI DELLA COMUNICAZIONE ESPRESSIVA DELLA MUSICA BASATA SU ESPERIMENTI PERCETTIVI

Laureando: Diego Villatora

Relatore: Giovanni De Poli

Correlatori: Sergio Canazza, Antonio Rodà

Corso di Laurea Magistrale in Ingegneria Informatica

Anno Accademico 2010-2011

Sommario

L'espressività è una delle componenti principali di una performance musicale. Svolge un ruolo importante nel rendere la musica interessante da ascoltare e ricca di emozioni. In questa tesi vengono analizzati dei lavori presenti in letteratura, derivanti da scuole di pensiero differenti. Inoltre viene affrontato il problema della definizione di una struttura adeguata per il riconoscimento delle emozioni della musica.

Sono proposti due esperimenti di tipo percettivo – con relative analisi e discussioni dei risultati – nei quali si cercano di chiarire alcuni aspetti riguardanti la dimensionalità dello spazio delle intenzioni espressive nella musica. Il primo esperimento ha lo scopo di indagare l'organizzazione delle emozioni a livello mentale degli ascoltatori, durante l'ascolto di una serie di brani del repertorio classico occidentale, quando viene a mancare una delle caratteristiche più importanti che aiutano nella discriminazione degli estratti, ossia la differenza tra modalità maggiore e minore. Nel secondo esperimento è posta una ulteriore restrizione alla selezione dei brani, con lo scopo di eliminare, oltre alla componente della diversa modalità di esecuzione, anche il differente metronomo; la scelta comporta brani in maggiore e con lo stesso tempo di esecuzione. Questi esperimenti percettivi limitano i fattori principali di discriminazione durante l'ascolto musicale, mettendo in evidenza fattori secondari indispensabili per la costruzione di un modello di riconoscimento delle emozioni nella musica. Tale struttura a bassa dimensionalità è di particolare interesse nel settore della *Sound and Music Computing*, con numerose possibilità di applicazione nei campi riguardanti: (a) il miglioramento della progettazione delle interfacce multisensoriali; (b) la mediazione tecnologica per l'accesso e la fruizione di contenuti musicali in funzione delle emozioni dell'utente, fondamentale per il *Music Information Retrieval* che si propone nuove soluzioni per il reperimento dei brani non solo attraverso campi predefiniti, come il titolo o l'autore, ma anche utilizzando le intenzioni espressive associate alla musica.

Indice

Capitolo 1

Introduzione

Nella vita di tutti i giorni ciascuno di noi è accompagnato per gran parte della giornata dalla musica, che essa sia ascoltata volontariamente o sia solo di sottofondo. è probabilmente vero, inoltre, che la maggior parte dell'esperienza musicale delle persone è spesso relazionata con una risposta affettiva di qualche tipo. Le emozioni possono essere identificabili come esperienze passivamente subite dal soggetto, sono solo in parte relazionate alle cause che le provocano; sono essenzialmente non cognitive.

Quando si parla di qualcosa che esprime un'emozione, solitamente implica il fatto che venga svelato pubblicamente lo stato provato in quel momento, che si possano vedere e percepire delle azioni o dei cambiamenti relativi al proprio stato affettivo; per esempio le lacrime di una persona sono relazionate all'emozione che noi definiamo tristezza, se il soggetto sta effettivamente provando uno stato emotivo triste. Le persone riescono in molti modi, per lo più fisiologici, a esprimere ciò che provano. Verranno esposti in dettaglio nel capitolo 2 i principali modelli strutturali per le emozioni ricavati da studi in ambito psicologico; queste strutture dimensionali sono importanti per determinare la percezione delle emozioni nelle persone.

La musica, al contrario, non esprime emozione secondo questa interpretazione, piuttosto viene recepita in maniera soggettiva da ciascuno di noi, e viene elaborata in maniera differente, a seconda del background sociale, culturale e di esperienze di vita che ciascuno porta con sè.

Parlare di percezione uditiva significa occuparsi dei suoi vari aspetti, e in particolare degli oggetti sonori su cui tale facoltà si esercita; rispetto alla capacità analitica di percezione, che appartiene a ognuno di noi, si può evidenziare la presenza di un'attenzione verso più aspetti riguardanti l'oggetto sonoro, alcuni più evidenti (intensità fortissimo, durate molto lunghe), che balzano immediatamente all'orecchio, e alcuni più sottili e sfumati (presenza di risonanze, cellule ritmiche o melodiche che si ripetono, leggere variazioni di intensità o di attacco del suono) (Imberty, 1986).

1.1 Prospettive psicologiche sull'espressività della musica

L'emozione è uno degli aspetti dell'esistenza umana più pervasivi, in relazione con quasi ogni aspetto del comportamento umano, come possono essere le azioni, le percezioni, la memoria, l'apprendimento e le decisioni che vengono prese (Sloboda e Juslin, 2001). Tra i primi a studiare, da un punto psicologico, le emozioni ci furono già a partire dal diciannovesimo secolo lo stesso Darwin (1872) e James (1884); seppur studiata in gran parte, la comprensione di come si sviluppano le emozioni è sempre stata messa in secondo piano rispetto a processi mentali di livello superiore, come il ragionamento e il problem solving.

La domanda: "Che cos'è un'emozione?" non è di facile risposta, e tutt'ora non vi è una sola scuola di pensiero; una delle ragioni è che le emozioni sono difficili da definire e da misurare. Russel esprime un concetto molto chiaro in un suo lavoro, cioè che "tutti sanno che cos'è un'emozione fino a che non viene loro chiesto di darne una definizione" (Russell e Fehr, 1984). La parola emozione è sia un concetto popolare, di vita quotidiana, e anche un costrutto scientifico; porta con sè sia una parte di significato esplicito che implicito. La parte implicita è incorporata nel concetto definito come "teorie popolari" dell'emozione; ossia ciascuno pensa di sapere che cos'è un'emozione, come questa si relazione allo stato affettivo di ciascuno, e viene nella maggior parte delle volte catalogata con termini anch'essi di tradizione popolare, come triste, felice, ecc. La parte di conoscenza esplicita si focalizza sullo studio di come l'emozione si sviluppa, che relazioni vi sono tra uno stato affettivo e i relativi cambiamenti fisiologici in una persona, ecc.

La costruzione dell'emozione, da un punta di vista scientifico, è dedotta da tre tipi di evidenza:

1. **Self − report**. La scelta più comune, e semplice, per valutare la risposta emotiva in una persona adulta è attraverso un'analisi auto-valutativa; questa può utilizzare tutta una seria di strumenti, come possono essere delle tabelle di aggettivi, *adjective checklist*, oppure delle scale di valutazione, o ancora dei questionari o descrizioni libere. Questo approccio ha associati alcuni problemi, tra i quali l'imperfezione di fondo tra le parole e l'emozione associata, e la scelta per esempio di quali parole comprendere nella lista di aggettivi. Allo stesso tempo però, questo risulta essere anche il metodo più diretto per valutare l'evidenza di uno stato emotivo, e certi stati emotivi non possono essere raggiunti se non con questa tecnica.

2. **Comportamento Espressivo**. non sempre la prima tecnica è possibile, e quindi si devono valutare altri aspetti , come possono essere la vocalizzazione, l'espressione facciale e il linguaggio del corpo. Anche questo tipo di approccio non è sempre accettabile, per il fatto che non sempre uno stato emotivo viene ben definito da questi parametri, e non sempre un comportamento espressivo porta con sè un effettivo stato emotivo, ma può essere

provocato volontariamente dal soggetto per comunicare informazioni agli altri.

3. **Misure Fisiologiche**. sfruttano la connessione esistente tra lo stato emotivo e un cambiamento a livello fisiologico interno del soggetto. Anche in questo caso però, seppur in forma minore che al punto due, è possibile la volontaria manipolazione di alcuni parametri anche in assenza di emozioni; ciò che è più importante però, è che non è stata stabilita finora una relazione abbastanza chiara tra stato emotivo e risposta fisiologica.

1.1.1 Modelli affettivi

Nel corso della storia dell'uomo vi sono state molte correnti di pensiero riguardanti la possibile strutturazione delle emozioni.

Darwin fu il primo che scientificamente cercò di esplorare questo campo, notando che tutta una serie di espressioni facciali e posizioni del corpo degli esseri umani sono simili a quelle di altri esseri umani e che quindi derivano dal lento e costante processo di evoluzione della razza. Secondo questo pensiero, l'evoluzione può essere considerata come primo aspetto sulla base del quale analizzare le varie emozioni. Ispirati a ciò molti ricercatori hanno sviluppato sistemi che attraverso il riconoscimento di espressioni facciali simili e/o posture particolari del corpo riescano a riconoscere la particolare emozione provata dall'utente.

Questo primo frame work è stato sviluppato in più occasioni da altri studiosi, anche perché sebbene la teoria darwiniana delle emozioni fallisca nella spiegazione di particolari comportamenti ed espressioni, tutta una serie di espressioni facciali di emozioni sono state riconosciute come basilari e quindi utilizzate.

Un altro modello, che in parte espande quello definito da Darwin, è la "teoria di James-Lange", che combina l'espressione con la fisiologia, e in cui la percezione di cambiamenti a livello fisiologico in un soggetto sono interpretati non come risultato di un diverso stato emotivo, ma piuttosto come emozione stessa. Questo pensiero sostiene che se cercassimo di astrarre dalla nostra coscienza tutto ciò che è sintomo fisico, allora non resterebbe più nulla di emotivo. Negli ultimi anni sono stati fatti notevoli passi in avanti nello studio di queste manifestazioni fisiologiche; in particolare lo sviluppo tecnologico ha permesso di avere a disposizione strumenti sempre più performanti nell'estrazione e nell'elaborazione di questi segnali, con un livello di informazioni ricavate maggiore.

Il monitoraggio di segnali fisiologici, come aspetto negativo, ha il fatto di essere in parte invadente per l'utilizzatore, anche se negli ultimi anni l'integrazione di dispositivi fisiologici per il monitoraggio è stata, con successo, resa meno intrusiva, permettendo di avere quindi un vasto database di dati su cui utilizzare metodi di statistical learning per il riconoscimento di path significativi da utilizzare per l'interpretazione dello stato emotivo.

L'approccio cognitivo, invece, definisce la possibilità di una persona di provare emozioni, data da una particolare interazione con un oggetto o da un evento, come una diretta conseguenza basata sull'esperienza della persona stessa, sullo scopo che ha in quel determinato momento e sull'azione che sta compiendo.

Quindi secondo la teoria cognitiva per cercare di predire come una persona reagirà a un determinato evento si devono conoscere sia gli obiettivi che le aspettative in quel momento della persona stessa.

Questo modello di definizione dell'emotività è stato analizzato e implementato in alcuni progetti di ricerca, con buoni risultati nella predizione di quale emozione l'utente stesse provando. Se però si pensa a un impiego di questa teoria in una situazione reale di per sè molto complessa, si riscontra facilmente il problema, se non l'impossibilità, di costruire la conoscenza di base necessaria per questo tipo di valutazione.

1.1.2 Approcci usati per l'analisi dell'emotività

Il modo in cui gli scienziati hanno concettualizzato l'emozione si differenzia in tre scuole di pensiero principali: il metodo categorico, il metodo dimensionale e il modello del prototipo.

Approccio Categorico Secondo questo metodo procedurale, le persone sperimentano le emozioni come categorie che risultano essere distinte tra loro. Essenziale per questo approccio risulta essere il concetto di "emozioni di base", le quali sono riconosciute come universali e considerate come concetti fondamentali a cui tutte le altre si riconducono in parte (Plutchik, 1994). Si cerca inoltre di dare una definizione funzionale in termini di raggiungimento di uno scopo per ciascuno delle categorie; per esempio la categoria felicità viene associata a un ragionevole progresso verso un'obiettivo (Oatley, 1992).

L'approccio categorico può essere particolarmente utile nei casi in cui viene richiesta una maggiore velocità di comprensione rispetto a un precisione migliore; algoritmi di riconoscimento delle emozioni possono agire in situazioni di tempo limitato, con conoscenza e capacità cognitive non complete.

Sono stati proposti in letteratura diversi criteri per la distinzione tra emozioni di base ed emozioni secondarie o complesse. Il metodo più importante classifica le emozioni di base per cui:

- hanno particolari funzioni che contribuiscono al retaggio dell'individuo
- sono riscontrabili in qualunque cultura
- sono sperimentabili come un'unico stato emotivo
- sono associate con distinti cambiamenti dei parametri fisiologici
- hanno espressioni emotive distinte

Le emozioni che sono definite secondarie sono comunemente pensate come una composizione di quelle di base, o particolari valutazioni cognitive che avvengono assieme a una emozione di base. Questo approccio è stato principalmente criticato per il fatto che non c'è una corrispondenza precisa tra i risultati dei vari ricercatori, ossia differenti gruppi di emozioni di base sono state presentate in letteratura; questa situazione può essere causata da una non coerente

definizione di che cosa definisce un'emozione, infatti se si valutano risultati che considerano lo stesso significato funzionale, si scopre un accettabile accordo tra quali debbano essere le principali emozioni, e almeno ne sono state riscontrate cinque: felicità, tristezza, rabbia, paura e disgusto (Oatley, 1992).

Approccio Dimensionale L'approccio categorico si focalizza principalmente sulle caratteristiche che distinguono un'emozione da un'altra. In contrasto, l'approccio dimensionale si concentra nell'identificare ogni emozione in una precisa posizione di una struttura dimensionale con poche dimensioni, solitamente due o tre. Una maggiore analisi nel dettaglio per questo tipo di approccio, che è stato utilizzato in questo lavoro, verrà trattata più avanti nella relazione.

Approccio Prototipo Per questo tipo di approccio si basa sull'idea che il linguaggio e le strutture associate a esso, riescano a dar forma su come le persone possano concettualizzare e classificare le informazioni (Rosch, 1978). Un principio base è che l'appartenenza a una particolare categoria è determinata da una somiglianza con l'esemplare prototipo di quella categoria. Di per sè il prototipo è un'immagine astratta che consiste in una serie di parametri che rappresentano l'esemplare tipico della famiglia.

Questo approccio fornisce un compromesso tra i primi due analizzati finora, perché mira ai contenuti delle categorie individuali e anche a una relazione gerarchica tra le categorie. Le critiche maggiori a questo approccio vedono il fatto che le informazioni ricavate da esperimenti su soggetti diversi spesso sono insufficienti a creare una struttura così articolata; inoltre alcuni critici mettono in risalto il fatto che non sempre vi è accordo tra quali emozioni debbano essere qualificate come prototipi di livello base e soprattutto non vi è una coerenza tra i confini delle strutture dei prototipi, con conseguente incertezza nel classificare alcuni stati emotivi.

1.1.3 Studio dell'emozione in relazione alla musica

Nella maggior parte dei trattati scientifici che si sono posti l'obiettivo di studiare le emozioni pochi sono i casi in cui viene approfondito l'argomento di reazione emotiva alla musica; questo può riflettere la sensazione che le emozioni percepite ascoltando musica siano differenti in un certo senso dalle altre. Quello che è vero è che ci sono importanti differenze tra le emozioni musicali e le altre sia riguardo agli antecedenti e sia alle conseguenze che provocano, ma questo non implica che siano emozioni diverse tra loro. Negli ultimi anni invece, la ricerca si è soffermata più volte sullo studio delle emozioni musicali, anche in conseguenza ad alcuni lavori di psicologi nel quale affermano che l'arte, nello specifico la musica, evoca forti reazioni emotive nonostante la sua natura illusoria.

Le emozioni percepite con la musica possono essere di due tipi:

- emozioni relative al valore estetico della musica; una risposta estetica alla musica è definita come un'intensa esperienza personale che coinvolge, tra le altre cose, componenti sociali, emotivi, cognitivi (Konecni, 1979).

- emozioni indotte o espresse dalla musica, più o meno legate al valore estetico della stessa; quest'ultime saranno anche le emozioni trattate in questa tesi.

Non vi è comunque una netta distinzione tra le due categorie, in particolare è difficile separare nettamente le une dalle altre perché in molti casi sono due facce della stessa medaglia. Una valutazione migliore delle seconde però, è possibile, infatti essendo più legate alla struttura musicale possono essere comprese e nei limiti imitate parzialmente; le prime al contrario, sono argomento più specifico per i ricercatori psicologi, in quanto la natura umana di ciascuno vede sfumature e particolari difficilmente quantificabili.

Lo studio delle emozioni nella musica è un eccellente mezzo per la comprensione del comportamento emotivo; l'uso di stimoli musicali per studiare le emozioni ha una validità di fondo in quanto le persone sono abituate a dare giudizi riguardanti la musica e a dare una valutazione alle loro reazioni affettive (Gaver e Mandler, 1987). Esperimenti condotti da Imberty hanno aiutato a comprendere la semantica con cui le persone esprimono le emozioni provate durante l'ascolto: lasciando totale libertà nella scelta degli aggettivi descrittivi, Imberty ha scoperto attraverso analisi fattoriali, che vi sono aggettivi usato più spesso di altri, termini che creano minor incertezza nella definizione rispetto ad altri che vengono interpretati diversamente da soggetto a soggetto (Imberty, 1986). Una selezione di questi aggettivi semanticamente rilevanti è stata utilizzata in molte analisi successive a Imberty, come nel caso degli studi condotti da Canazza e presentati nella sezione 3.2.3.

La musica inoltre, ha una struttura ricca e ben conosciuta, e la comprensione della struttura di un evento può aiutare in particolare a capire la reazione emotiva che esso comporta.

1.2 Protocollo sperimentale

La simultanea influenza di diversi spunti musicali rispetto alla percezione dell'ascoltatore riguardo alla percezione della similitudine tra estratti musicali deve essere valutata e gestita attraverso una corretta metodologia di raccolta dei dati e della successiva analisi. Si contrappongono due metodi per la risoluzione di questo punto, l'approccio cosiddetto "riduzionista" e l'approccio olistico.

L'approccio riduzionista comporta una serie di esperimenti controllati per l'estrazione dei risultati da varie dimensionalità musicali, e integra la collezione delle informazioni su dimensioni indipendenti in un modello unico e globale. L'approccio olistico, invece, è basato su esperimenti con un grande numero di stimoli, non limitato a pochi stimoli di base, per osservare la complessa relazione che avviene tra alcune variabili in analisi; quest'ultimo approccio, nel quale non sono ancora state definite le dimensioni principali della struttura a bassa dimensionalità che rappresenta le variabili, necessita di tutte le informazioni riguardanti il segnale dell'estratto, la completa forma d'onda.

Nel caso in cui si presenti la necessità di indagare le proprietà di una complessa struttura multidimensionale, come possono essere il timbro, il genere, o

l'intenzione espressiva - caso specifico di questa dissertazione - l'approccio tipico cerca di valutare la musica olisticamente per identificare le dimensioni nella musica mentre si stima la loro rilevanza. Il vantaggio di usare un approccio di tipo olistico vede la possibilità di relazionare direttamente al materiale originale utilizzato nell'esperimento le eventuali dimensioni utilizzate, come avviene in una normale elaborazione di percezione umana.

La difficoltà che comporta, però, l'uso di un metodo olistico è il fatto di maneggiare una selezione di stimoli complessa, con features che legano tra loro in maniera talvolta difficile da comprendere. Altro aspetto di difficoltà per tale approccio è il fatto di isolare e valutare correttamente l'effetto di una specifica variabile sulla rispettiva dimensione musicale, in rispetto al giudizio fornito dai soggetti partecipanti all'esperimento.

Per questa tesi si è scelto di seguire una metodologia olistica, in quanto lo scopo finale è la maggiore comprensione dell'effetto di alcune variabili sulla dimensionalità dello spazio delle intenzioni espressive nella musica; inoltre è scopo di questa tesi anche l'analisi approfondita della struttura multidimensionale dell'espressività musicale, per evidenziare, nel caso dovessero presentarsi, ulteriori dimensioni definite in aggiunta allo spazio bidimensionale principale trattato in letteratura, e analizzato nei capitoli due e tre di questa tesi.

1.3 Valutazione sperimentale della similitudine nella musica

Non esiste un'unica metodologia per la misura tramite sperimentazioni della similitudine tra brani musicali. In letteratura tre metodi sono citati, compatibilmente con gli scopi di questa tesi, e sono: *pair-rating, pair-ranking* e l'*object-grouping* (MacRae et al., 1990). Mentre i primi due metodi vedono la classificazione o valutazione su di una scala di valori delle varie coppie di estratti comparate, il metodo del raggruppamento da la possibilità ai partecipanti di raggruppare secondo la propria percezione i brani proposti, senza limiti di numero di gruppi o numerosità degli stessi. Quest'ultimo metodo vede inoltre una procedura di esecuzione più semplice, e più intuitiva da parte dei soggetti, i quali potrebbero avere invece problemi di valutazione della scala di valori su cui valutare le coppie secondo le altre due metodologie.

Unico problema dell'approccio per raggruppamento è la difficoltà a memorizzare le percezioni da parte dei soggetti se dovessero esserci troppi brani da valutare. Per evitare questa evenienza, negli esperimenti proposti in questa tesi, il numero di brani non è mai stato troppo elevato, una ventina circa per ogni esperimento, e la loro lunghezza in media circa una ventina di secondi; inoltre i partecipanti hanno avuto completa libertà nell'ascolto dei brani, sia come quantità di ascolti, sia nell'ordine di ascolto, e sia come durata dell'esperimento stesso.

1.4 Struttura della Tesi

In questo capitolo sono stati introdotti i concetti principali inerenti agli scopi di questa tesi; sono stati presentati, inoltre, vari metodi di approccio allo studio dell'informazione espressive della musica, e sono state evidenziate alcune scelte metodologiche utilizzate per gli esperimenti proposti. Nel capitolo 2 verranno messi a confronto i due maggiori modelli a livello psicologico per l'organizzazione mentale delle emozioni, ossia il modello a due dimensioni, tipico della scuola nord-americana, e il modello tridimensionale supportato dalla ricerca europea. Nel capitolo 3 viene introdotto il settore dell'*Emotional Computing*, con le differenze dovute a scuole di pensiero diverso e con l'illustrazione di alcuni esempi applicativi; in particolare viene presentato il *Kinestetic space* relativo alla struttura di percezione sensoriale della musica. Nel capitolo 4 vengono illustrati due lavori che descrivono lo stato dell'arte per quanto riguarda i due approcci più importanti per lo studio dell'espressività musicale: analisi di tipo *score-dependent* e analisi *score-independent.* Nei capitoli 5 e 6 vengono esposti i due esperimenti proposti in questa tesi, e vengono discussi i risultati ottenuti. Infine, nel capitolo 7, vengono riassunte le informazioni ricavate e vengono illustrati sviluppi futuri inerenti a questo lavoro.

Capitolo 2

Modelli psicologici per l'emozione

In questo capitolo vengono presentati e analizzati diversi risultati presenti in letteratura, derivanti da anni di ricerca a livello psicologico per quanto riguarda l'emozione e la sua struttura. Negli ultimi anni sono state condotte molte analisi ed esperimenti; l'interesse in questo campo non è una novità, come visto nel capitolo 1; il tema riguardante la struttura dell'emozione è uno tra gli argomenti più intensamente studiati, soprattutto in campo psicologico, e non sempre trova conformità di risultati e modelli condivisi.

Nella prima parte verrà spiegata la struttura a due dimensioni, tra cui i due risultati più importanti derivanti dai lavori di Russell (1980) e Thayer (1989). Nella seconda parte, invece, viene preso in considerazione un'approccio tipico della comunità europea, cioè un definizione strutturale a tre dimensioni. I vari modelli verranno poi relazionati tra loro e valutati rispetto a risultati consolidati derivanti da alcuni studi.

2.1 Struttura a due dimensioni per l'emozione

Come cita Russell in uno dei suoi lavori (Russell e Feldman Barrett, 1999), in soli sette anni, cioè dal 1991 al 1997, il *Journal of Personality and Social Psychology* ha pubblicato 359 articoli nei quali l'emozione era tra le variabili discusse, e questi saggi rappresentano il 29% degli articoli dell'intero periodo. Questo per sottolineare come la discussione relativa alla definizione e alla struttura delle emozioni sia un argomento di notevole interesse, ma con molte visioni differenti, talvolta contrastanti tra loro.

I confini del dominio dell'emozione sono così vaghi che a volte sembra che tutto possa far parte del termine emozione (Russell e Feldman Barrett, 1999). In questa sezione verrà presentata la struttura definita da Russell nel corso degli anni, struttura che in molti studi è stata ritenuta la più affidabile e autorevole, soprattutto per quanto riguarda i ricercatori americani.

2.1.1 Modello *valence-arousal*

Russell, negli ultimi anni, ha cercato di mettere insieme le idee e i risultati dei suoi lavori, definendo una propria idea di come debba essere interpretata l'affettività (Russell, 2003). In primo luogo, la distinzione più importante che viene fatta determina due concetti distinti, che sono il *core affect* e i *prototypical emotional episodes.*

- **Prototypical emotional episodes**. è un complesso insieme di sotto-eventi che riguardano uno specifico oggetto; quest'ultimo può essere una persona, una condizione, un evento o una cosa, reale o immaginaria che sia. Questo concetto è usato il più delle volte per indicare i casi più chiari di emotività, situazioni alle quali solitamente fa riferimento la gente comune. Questi episodi emotivi necessariamente includono anche il *core affect*, ma non solo; in parte sono costituiti dalla reazione alla particolare situazione creata in relazione con l'oggetto, l'esperienza che il particolare soggetto ha riguardo a tale emotività, e infine tutti gli aspetti di carattere neurale, chimico e corporeo che si scatenano come conseguenza a tale evento. Per il fatto che questi episodi sono connessi direttamente a un oggetto, hanno una forte componente di carattere cognitivo, sia di processo che strutturale. Infine, come dice la parola stessa "episodio", sono situazioni momentanee, che hanno un inizio e una fine, cioè avvengono in una determinato intervallo temporale che è possibile quantificare.

- **Core affect**. Con questo concetto Russell si riferisce alla parte di affettività più elementare e accessibile, e a tutte le conseguenze neurofisiologiche che comporta; un sentimentalismo affettivo che non è rivolto a un particolare oggetto, com'era il caso per gli episodi emozionali. Il *core affect* si protrae nel tempo, può essere interpretato come libero stato emotivo oppure entrare a far parte di un evento più significativo com'è un episodio emozionale. In qualunque momento, comunque, anche questa componente è sempre causata da particolari fattori, come possono essere il tempo o i cicli diurni. Il *core affect* varia in intensità a seconda del momento, ma ciò che è importante è che chiunque si trova sempre in un qualche stato di *core affect*, sia uno stato relativo ad un emozione definita come per esempio felicità o sia solamente uno stato neutro.

I concetti presentati sopra quindi, seppur relazionati tra loro, sono distinti; un episodio emozionale può essere sezionato in più componenti, e il *core affect* è una di queste, ed è la parte più importante di qualsiasi episodio o situazione provata a livello emotivo. Un esempio di come può essere identificato il *core affect* è lo stato d'animo presente alla mattina quando ci si sveglia, che può essere tra i più vari, e in pochi casi si riesce a identificare un motivo che lo possa aver causato; stesso dicasi per una situazione di malessere che influisce pesantemente sullo stato d'animo del soggetto ammalato.

2.1.1.1 Struttura dell'emozione: il *circumplex*

Russell (2003) prende in considerazione tre approcci che riassumono quasi ogni metodologia usata in letteratura per ricercare una struttura adatta a descrivere i *prototypical emotional episodes* e il *core affect*.

L'approccio usato da molti è di tipo categorico, ossia il tentativo di riuscire a identificare una serie di categorie di base, mutuamente esclusive, le quali siano in grado di descrivere anche tutte le altre categorie derivanti dall'interazione delle prime. I risultati derivanti da analisi anche di tipo diverso sono discordanti, non convergono a un insieme chiaro e definito di categorie basilari. Inoltre a questo, se si considerano linguaggi diversi si distinguono un numero di categorie diverso; per esempio, in inglese sono state riscontrate dalle 500 alle 2000 categorie, mentre in Ifaluk ve ne sono 50 e in Chewong solamente 7.

L'approccio gerarchico è stato usato in letteratura per catturare la relazione di sotto-tipo di alcune categorie rispetto ad altre. Anche questa metodologia è stata criticata da Russell per il fatto che l'emozione è un entità astratta e con confini non precisi, e quindi è di difficile interpretazione tutta una serie di relazioni gerarchiche di composizione tra categorie o di mutua esclusione tra concetti allo stesso livello di gerarchia.

Russell suggerisce quindi una struttura dimensionale. Gli episodi emozionali variano lungo due dimensioni che sono rappresentate dal grado di *pleasure* e dal grado di *activation* provata. Ogni situazione ha però una differente caratterizzazione nel piano, per esempio episodi emotivi di paura differenti possono identificare valori di *pleasure* differenti. L'analisi fattoriale di emozioni raccolte dall'autore e dai suoi collaboratori tramite analisi auto-valutative, associata a un'analisi multidimensionale ha portato a questa conclusione, dove le due dimensioni sono state nominate *pleasure* e *arousal*. Le categorie emotive non si clusterizzano lungo gli assi di questo spazio: la struttura risultante è il cosiddetto *circumplex* (Russell, 1980). Sebbene Russell sia stato tra i più importanti sostenitori di questo risultato, lo stesso ammette che questa struttura dimensionale poco si adatta a specificare l'intero insieme degli episodi emozionali, che anzi cadrebbero solo in alcune parti del grafico (Figura 2.1); la struttura a due dimensioni sarebbe invece limitata alla sola rappresentazione del *core affect*.

Lo stesso autore ammette come la suddetta struttura bidimensionale non riesca a distinguere tra episodi come paura, rabbia, imbarazzo e disgusto, che condividono lo stesso *core affect* e quindi cadono nello stesso luogo nel *circumplex*. Quindi il modello *pleasure-arousal* rappresenta una sola componente dell'episodio emozionale, ma non le altre specifiche, le quali permettono poi di riuscire a specificare episodi come quelli citati in precedenza. Una nota da fare è il fatto di aver individuato un *circumplex* invece di una struttura semplice; questo viene spiegato con la definizione di entrambe: la prima prevede che la variabili al suo interno siano intercorrelate in maniera da rappresentare un cerchio, o come minimo abbiano distanza simile dall'interno dello stesso, anche non avendo una spaziatura equa; la seconda invece risulta nel caso in cui le variabili siano correlate con solo uno dei fattori della struttura, cosa che non avviene in questo caso.

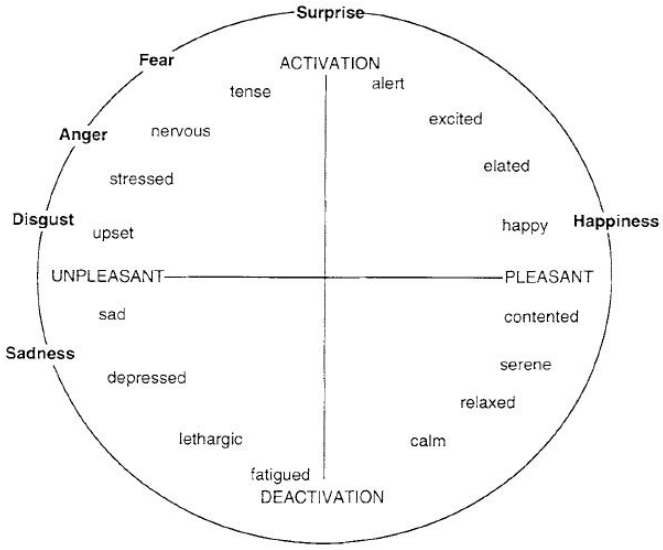


Figura 2.1: *Circumplex* di Russell, con all'interno la struttura del *Core Affect* e all'esterno alcuni episodi emozionali rappresentati. (Russell e Feldman Barrett, 1999)

2.1.1.2 Altri modelli a confronto

Per confermare la propria struttura Russell considera alcuni modelli bi-dimensionali presi dai lavori più autorevoli che propongono modelli con la medesima dimensionalità, in particolare il modello a due dimensioni dell'*arousal* di Thayer (1989), vedi sezione 2.1.2, il risultato dello studio di Watson e Tellegen (1985) e il lavoro di Larsen e Diener (1992). A questi modelli Russell accosta il proprio, e ne valuta la capacità di rappresentare le dimensioni proposte negli altri. I risultati confermano che il modello di Russell comporta una compatibilità che varia dal 73% al 97% con le altre strutture, fatto che viene spiegato con l'idea che tutti questi modelli si basano su una base comune, e che le dimensioni identificate sono tipicamente le stesse, a meno di errori di misurazione o di interpretazione dei risultati, e naturalmente dal nome date a esse.

Sono state commentate, inoltre, strutture dimensionali diverse da quella bi-dimensionale presentata. Le soluzioni che vedono una sola dimensione sono ritenute troppo difficili da interpretare, sia dal punto di vista della loro denominazione, sia dal fatto che la perdita di informazione nel modello unidimensionale è eccessiva e non accettabile. Prendendo in considerazione invece le soluzioni alternative che identificano più dimensioni, Russell cita alcuni risultati in cui sono state trovate dimensioni aggiuntive, denominate di volta in volta come potenza, dominanza, dovute alla causa. Russell interpreta queste dimensioni come ausiliarie dell'evento che scatena la reazione e quindi le esclude dalla sua definizione di *core affect*. Queste ulteriori fattori latenti non vengono considerati da

Russell come un miglioramento del modello, seppur in forma minoritaria, ma vengono attribuiti a cause di altro tipo, non andando a intaccare il modello di *pleasure-arousal*.

E forse questa è la questione aperta su cui molte teorie si vedono in contrasto, ossia sul fatto della dimensionalità dello spazio emotivo. Non volendo modificare o estendere la propria struttura, Russell limita la funzione di questi ulteriori possibili fattori a effetti complementari dell'episodio emozionale, senza per questo mettere in discussione lo spazio del *core affect.*

2.1.2 Thayer - La biopsicologia dell'umore e dell'*arousal*

Il concetto di *arousal* è per molti aspetti di difficile definizione, perché riguarda una serie di processi umani di per sè sono molto complessi. Ciò non toglie il fatto che questo fattore sia da tutti gli studiosi riconosciuto e relazionato, in maniere differenti a seconda della teoria, a particolari processi biologici e di scienze comportamentali. Nella vita di tutti i giorni passiamo da stati di calma a momenti di attivazione maggiore, e viceversa, quindi un *continuum* di cambiamenti biologici e non solo avviene in continuazione, per esempio uno stato di sedentarietà in contrapposizione con un momento di attività fisica, oppure una intensa emozione.

Il concetto di *arousal*, rispetto al pensiero di Thayer, è che sia un sistema con una struttura di fondo multidimensionale, nello specifico un'attivazione di tipo energetico e una di tensione (Thayer, 1989).

2.1.2.1 Struttura dell'*arousal*

Per studiare la dimensionalità dell'*arousal* attraverso l'uso del linguaggio, è stata preparata una lista di aggettivi in un formato per cui i partecipanti al test fossero in grado di definire per ogni aggettivo la loro momentanea situazione emotiva rispetto allo stesso, usando una scala a quattro valori. Siccome il presupposto dell'analisi di Thayer era definire una struttura a più fattori per l'attivazione, il test è stato riproposto più volte nell'arco della giornata e in condizioni differenti tra i partecipanti, ossia in situazioni di tranquillità o dopo uno sforzo fisico, per avere un quadro generale sul proprio auto-giudizio che i partecipanti davano rispetto ai vari termini proposti.

Dopo aver collezionato una notevole serie di dati, Thayer ha analizzato ripetutamente i valori ricavati, tramite analisi fattoriale, per ricavare il numero minimo di fattori latenti in grado di rappresentare questo gruppo elevato di variabili. Sono emersi, dall'analisi, quattro fattori distinti; sulla base di queste categorie sono stati raggruppati tutti gli aggettivi usati:

- *General Activation* o *Energy*, con aggettivi come energetico, attivo, vigoroso
- *Deactivation-Sleep* o *Tiredness*, rappresentato da termini come insonnolito, stanco, addormentato

- *High Activation* o *Tension*, rappresentante teso, spaventato, intenso
- *General Deactivation* o *Calmness*, in rappresentanza di stati emotivi come calmo, pacifico, placido

Analisi successive, con tecniche di rotazione obliqua e analisi del secondo ordine, hanno identificato un importante risultato, ossia la correlazione negativa tra i primi due fattori elencati sopra, e questo ha portato a identificarli come due poli di una stessa dimensione, l'*energetic arousal*; inoltre la stessa analisi ha sottolineato anche la correlazione negativa tra gli ultimi due fattori dell'elenco, e quindi anch'essi sono stati interpretati come parti opposte di uno stesso asse, etichettato come *tense arousal*.

Thayer ha quindi proposto che la struttura principale su cui si evolve l'*arousal* non è unidimensionale lungo un'unica dimensione di intensità, ma bensì bidimensionale lungo due fattori che interagiscono tra loro. La dimensione definita come energetica è relazionata con uno stato di energia, vigore contrapposto a uno stato di stanchezza, fatica e sonnolenza, mentre la dimensione di tensione varia da uno stato di tensione, paura fino ad arrivare a sensazioni di calma e quiete.

Lo stesso autore suggerisce il fatto di una possibile presenza di errori nella misurazione, dovuta a varie cause, sia di metodologia e di analisi che di situazione in cui è avvenuta l'autovalutazione; in particolare afferma che la covarianza tra le due dimensioni principali non sempre segue l'andamento ottimale per la distinzione di due fattori tra loro completamente ortogonali. Conclude supponendo la presenza di un'ulteriore distinzione tra i fattori risultanti, ossia i quattro di base citati sopra, che comunque non andrebbe a intaccare la bontà del modello specificato, ma sarebbe una utile specificazione dello stesso.

2.1.2.2 Altri modelli a confronto

Importante sottolineare come Thayer dia una personale comparazione tra il modello da lui proposto e altri modelli precedentemente presentati in letteratura, tra cui la più citata struttura bidimensionale di Watson e Tellegen (1985). Nell'analisi fatta, Thayer individua come differenza principale tra i due modelli la denominazione dei fattori, che nel modello di Watson e Tellegen si presentano col nome di affettività positiva e negativa. L'autore conclude questa osservazione indicando il valore affettivo che appartiene alle dimensioni da lui definite, con una forte relazione tra *tense arousal* e tono affettivo negativo, e tra *energetic arousal* e tono affettivo positivo.

Soprattutto quest'ultima analisi è stata utilizzata in letteratura più volte, successivamente al lavoro di Thayer, per spiegare come il modello bidimensionale dell'*arousal* riesca a riprodurre in parte la dimensione di valenza presente in molti altri modelli, a partire da Russell, e quindi per raffrontare le strutture dell'affettività.

L'*energetic arousal* appare come un sistema di attivazione generale, che comprende anche importanti segnali funzionali per la propria valutazione e per le

azioni che prevedono la necessità di prendere delle decisioni. Sono osservabili variazioni in questo primo sistema in relazione al ciclo diurno o circadiano, all'attività motoria volontaria e al periodo di riposo; anche condizioni fisiche e di salute influiscono particolarmente su questo fattore. Per quanto riguarda il sistema del *tense arousal*, quest'ultimo è relazionato con una forte componente cognitiva collegata a stimoli di pericolo e a reazioni biopsicologiche.

2.2 Struttura a tre dimensioni per l'emozione

Cercare di dare una struttura ben definita all'affettività è sempre stato uno dei principali settori di studio per quanto riguarda la psicologia. In particolare la discussione su quante debbano essere le dimensioni principali in un modello dimensionale è sempre stata accesa, a partire dalle argomentazioni di Wundt di fine ottocento fino ad arrivare ai giorni nostri. Wundt sosteneva il suo modello tridimensionale per le esperienze emotive; propose che le emozioni fossero una combinazione di sei emozioni basilari, a coppie tra loro mutuamente esclusive, ed erano denominate come *pleasure-displeasure, tension-relaxation e arousal-calmness.* In contrasto molti altri autori sostenevano tesi riguardanti la dimensionalità dello spazio emotivo con due o solo una dimensione.

Attualmente la discussione sembra non essere arrivata a un punto comune. La letteratura nord americana è favorevole a un modello bidimensionale, come visto nella sezione precedente, mentre i ricercatori europei sostengono una struttura a 3 dimensioni. Nelle prossime pagine saranno presentati i lavori più significativi riguardanti questo tipo di modello.

2.2.1 Considerazioni e definizione di un modello

Non sempre vi è stata una valutazione alla pari dei suddetti modelli, infatti in molte tra le pubblicazioni di ricercatori psicologi più famose negli scorsi anni non viene nemmeno citata la possibilità di ampliamento del modello bidimensionale a tre dimensioni. Per fare un esempio esplicativo citato in (Schimmack e Grob, 2000), tre delle maggiori pubblicazioni che supportano il modello tridimensionale sono state citate in letteratura 198 volte (Sjöberg et al., 1979), 290 volte (Matthews et al., 1990), e 76 volte (Steyer et al., 1994); in contrasto le maggiori pubblicazioni supportanti il modello a due dimensioni sono state citate 2140 volte (Watson e Tellegen, 1985), 2576 volte (Russell, 1980), e 647 volte (Thayer, 1989)[1].

Questa diversa interpretazione strutturale può derivare da una differente metodologia utilizzata; (Schimmack e Grob, 2000) suggerisce che il modello tridimensionale europeo sia trovato con un'analisi fattoriale in cui si sfrutta una rotazione obliqua, invece i ricercatori americani preferiscono modelli con fattori ortogonali tra loro. In Figura 2.2 viene proposto il modello tridimensionale, in maniera semplicistica, su due dimensioni solamente; questo però comporta la

[1] I numeri delle citazioni sono stati presi sfruttando Google Scholar, numeri aggiornati a marzo 2010.

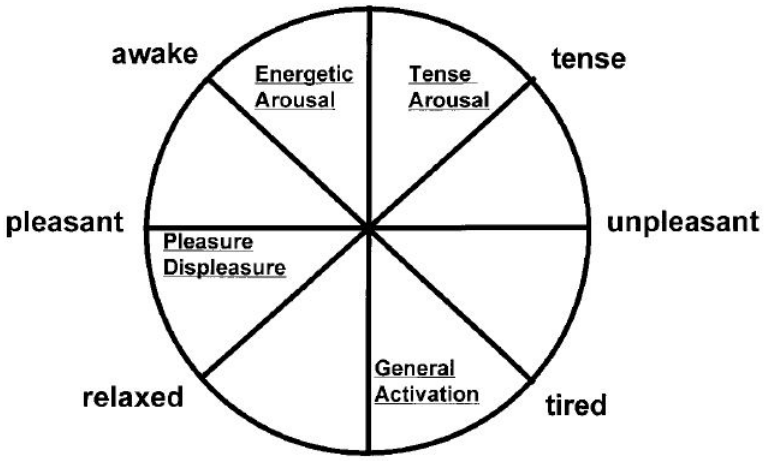


Figura 2.2: Modello Dimensionale del Core Affect. (Schimmack e Grob, 2000)

perdita di informazione e un minore adeguatezza a rappresentare i dati, come vedremo nel seguito.

Implicitamente i modelli con sole due dimensioni suggeriscono che ci si può ricondurre da un modello con più dimensioni a loro per il fatto che alcune dimensioni sono una composizione delle due basilari. Per esempio, nel modello di Thayer (1989), si suggerisce che l'asse relativo al *pleasure-unpleasure* e l'asse accostato all'*activation* siano piuttosto forme composte dei due assi ortogonali ricavati dall'autore, ossia una composizione di *tense arousal* e *energetic arousal*.

Si devono considerare però alcuni problemi che sono stati riscontrati utilizzando modelli bidimensionali:

- il problema principale di un modello a due dimensioni è che non si adatta completamente ai dati; anche autori come Russell e Watson ammettono che per avere una completa descrizione delle esperienze affettive sono necessari più dimensioni. Interessante citare il lavoro di Reinsenzein (1994), il quale nota che il modello bidimensionale non è in grado di distinguere tra specifiche emozioni come rabbia, invidia, disappunto e amore, gratitudine, orgoglio; lo stesso autore sottolineò come informazioni aggiuntive al modello a due dimensioni fossero necessarie per una migliore distinzione. I modelli con due dimensioni comportano che la correlazione tra i descrittori che formano il modello formi uno spazio bidimensionale; studi empirici mostrano però che questo tipo di soluzione non è adeguata a rappresentare l'intercorrelazione che esiste tra i vari descrittori del *core affect* (Watson et al., 1999). Schimmack afferma che il *tense* e l'*energetic arousal* siano dimensioni basiche, che non possano essere ridotte come composizione di *pleasure-displeasure* e una dimensione ortogonale che è l'*activity*; similmente, si afferma anche che la dimensione della *valence* non sia una mistura dei due *arousal*.

- Un altro problema riscontrato in letteratura è la scarsa definizione del concetto di *arousal*. Russel definisce la dimensione dell'*arousal* come "un continuo unidimensionale, che può essere descritto, a partire dalla fascia bassa, da sonno e sonnolenza, rilassamento, vigile, attivazione, iperattivazione, e per finire eccitazione frenetica, nella parte opposta" (Russell e Feldman Barrett, 1999). Questa definizione da parte di Russell comporta l'idea che l'*arousal* sia distribuito lungo una sola dimensione, fatto che non viene condiviso da altri autori, primo tra tutti Thayer, il quale parla di due dimensioni separate e ampiamente indipendenti di *arousal* descritte da sveglio-stanco o *arousal* energetico, e da teso-rilassato o *arousal* di tensione (Thayer, 1989).

  Una revisione del modello *pleasure-arousal* (Russell e Feldman Barrett, 1999) afferma che esiste una terza dimensione, ortogonale alle prime due, chiamata dimensione di attivazione generale; quello però che lo stesso Russell afferma, è che comunque vi siano delle composizioni tra le diverse dimensioni, ricavando difatti un modello sempre bidimensionale. Questo modello vede però nuovi problemi, tra tutti la difficoltà di trovare termini che descrivano la dimensione di attivazione; come sottolineato da Schimmack, vengono utilizzati descrittori che non possono far parte dei descrittori di base per l'affettività.

Il modello tridimensionale fornisce una chiara soluzione alla concettualizzazione dell'*arousal*. Per prima cosa questo modello riesce a riconoscere le separate dimensioni emerse dalla studio di Thayer, con due differenti tipi di *arousal*; in secondo luogo non viene definita una terza dimensione al modello bidimensionale che manca di descrittori nel linguaggio quotidiano; infine non viene fatta l'assunzione che l'*arousal* sia intrinsecamente piacevole o spiacevole, infatti anche intuitivamente si riesce a immaginare uno stato in cui ci si sente stanchi ma soddisfatti, per esempio dopo aver portato a termine un lavoro, e questo si scontra con la definizione data del modello *pleasure-arousal*.

Schimmack ha proposto dei test per verificare se la struttura del *Core Affect* sia bidimensionale. Vari modelli sono stati testati e ripetutamente hanno fallito nella conferma di questa supposizione. Il modello di partenza su cui queste analisi sono state fatte è tridimensionale, con la denominazione degli assi lasciata ad aggettivi bipolari, che sono: *pleasure-displeasure, awake-tiredness* e *tension-relaxation.* Per primo sono stati combinate *awake-tired* e *tense-relaxed* in un'unica dimensione di attivazione; questa soluzione male si è adeguata ai dati in quanto entrambe le dimensioni combinate sono negativamente correlate tra loro; chiaramente un individuo può essere sveglio e teso oppure sveglio e rilassato, e quindi l'*arousal* non è un costrutto unidimensionanle. Come secondo test Schimmack ha provato a spiegare la dimensione di *pleasure-displeasure* come combinazione dei due tipi di *arousal*; questo modello si è rilevato migliore del precedente, ma l'analisi degli indici di bontà del modello hanno indicato che deve essere scartato ugualmente al primo; importante sottolineare come sia stato provato che questo modello trascura circa il 50% della varianza dell'asse *pleasure-displeasure.* Infine, è stato testato il modello *pleasure-arousal*; e

neanche questo modello ha potuto dare un migliore adattamento ai dati, infatti ulteriori analisi hanno portato a scoprire che solo il 26% della varianza della dimensione *awake-tiredness* viene predetta. Il modello quindi, analizzato da Schimmack, da lui denominato PAT dalle iniziali delle etichette dei tre assi, si adatta bene per tutta una serie di test eseguiti, in confronto a un modello a due dimensioni.

2.2.2 Struttura dell'*arousal*

Nel 1989 Thayer propose due tipi differenti di attivazione, adattati al suo modello bidimensionale per strutturare lo stato affettivo umano, e sono l'*energetic arousal* e il *tense arousal.* Questa visione è stata più volte messa in discussione, e forse il punto principale su cui si basano le critiche è il fatto che molti studiosi hanno cercato di dare un'interpretazione ai due tipi di *arousal* come composizione delle due dimensioni base riconosciute da Russell, ossia *valence* e *activation.*

Già nel 1967, Thayer condusse uno studio sulla dimensionalità dell'esperienza di "attivazione" provata dalle persone, e quello che inizialmente ottenne furono quattro fattori indipendenti, due fattori legati all'attivazione e due fattori legati alla de-attivazione. In accordo con questi risultati, il concetto di attivazione è stato strutturato in maniera bidimensionale, variante lungo la dimensione definita con il nome di *energetic arousal*, che definendola con aggettivi emotivi varia da sentirsi assonnati a sentirsi svegli, e lungo la dimensione del *tense arousal*, che varia da uno stato di calma a uno stato in cui ci si sente nervosi (Thayer, 1989).

Questa concettualizzazione fatta da Thayer è supportata da diversi punti importanti. Primo tra tutti il fatto che le due dimensioni definite siano relazionate a cause differenti; per esempio, l'*energetic arousal* è influenzato dal ritmo circadiano, che corrisponde all'attività delle cellule del cervello che regolano il ciclo dell'organismo per quanto riguarda gli stati transitori da dormiente a sveglio; invece questa relazione non esiste con la seconda dimensione che vede una scala di stati relativi alla tensione. Come secondo punto è stato provato con alcuni esperimenti che queste due dimensioni di attivazione possono cambiare in direzioni opposte; l'esempio qui riportato è un esperimento in cui è stata analizzata l'influenza di una controllata ipoglicemia in alcuni pazienti, ed è stato riscontrato che laddove il livello di *energetic arousal* decresce come conseguenza di un basso livello di zucchero nel sangue, invece si nota un aumento del *tense arousal*, causato probabilmente da una risposta di emergenza del corpo che richiede un'azione immediata per ristabilire i livelli di zucchero nel sangue. Infine, come terzo punto, è stato provato in alcuni studi che i due tipi di *arousal* hanno conseguenze diverse, per esempio l'*arousal* legato allo stato di tensione è meno predittore di operazioni cognitive rispetto al secondo tipo di attivazione.

Successivamente al lavoro di Thayer, diversi articoli hanno messo in discussione questa visione separata dei due tipi di *arousal*, come descritto in precedenza, basandosi sul fatto che queste due dimensioni siano una composizione di un'unica dimensione di attivazione e della più riconosciuta dimensione della

valenza. Sentirsi svegli descrive uno stato di alta attivazione e di valenza positiva, laddove sentirsi stanchi descrive uno stato con bassa attivazione e valenza negativa. Similmente, sentirsi tesi è in relazione a un'alto livello di attivazione con valenza negativa, mentre sentirsi calmi descrive un basso livello di attivazione e una valenza positiva. Intuitivamente, già questa rappresentazione delle dimensioni non è del tutto corretta; bastano un paio di esempi per trovare dei casi che non cadono nella descrizione appena fatta: per esempio il sentirsi stanchi dopo aver portato a termine con successo un'operazione avrà comunque un basso livello di attivazione, e quindi energetico, però avrà una valenza catalogabile come positiva; altro esempio è il fatto di ritrovarsi svegli quando si cerca di addormentarsi non riuscendoci, stato in cui l'attivazione è alta, ma la valenza di certo non può essere descritta come positiva, soprattutto se all'indomani ci si deve svegliare presto.

Schimmack e i suoi collaboratori hanno proposto un test per verificare la natura di questi due tipi di *arousal* (Schimmack e Reinsenzein, 2002). Per condurre questo test è stato necessario misurare la correlazione tra i residui dei due tipi di attivazione in seguito alla rimozione della varianza attribuita, a entrambi, dalla componente della *valence*; quest'ultima è misurabile attraverso una scala valida e affidabile. In accordo con la teoria che afferma la struttura definita da valenza e attivazione, una volta rimossa la componente relativa alla valenza, il residuo restante dovrebbe rappresentare la varianza della componente attivazione; quindi secondo suddetta teoria, i residui dei due *arousal*, una volta rimossa la componente di valenza, dovrebbero riflettere la varianza lungo la stessa componente di attivazione, e quindi parlando in termini statistici, dovrebbero essere altamente correlati tra loro. In contrasto la teoria che afferma l'indipendenza dei due tipi di attivazione, vedi Thayer sezione 2.1.2, in relazione a questo tipo di test, dovrebbe comportare un indice di correlazione piuttosto basso tra i due residui di attivazione.

In alcune ricerche, ciò che ha distorto in maniera sostanziosa il risultato dei test è stato l'errore di misura. Per l'analisi fatta da Schimmack, sono stati valutati sia errori di tipo casuale e sia errori sistematici. Per quanto riguarda gli errori di tipo casuale nelle misure, sono stati controllati esaminando la relazione tra i costrutti a livello di fattori latenti. Invece, errori di tipo sistematico, sono stati limitati in primo luogo dalla valutazione su scala unipolare usata per i due tipi di *arousal* e poi sottraendo un polo dall'altro, e in secondo luogo per il fatto che gli errori sistematici, per loro stessa natura, sono condivisi tra la componente di attivazione e valenza, e quindi divengono parte della varianza condivisa tra attivazione e valenza, e non parte dei residui dell'attivazione.

La novità, nello studio presentato da Schimmack, sta nel non utilizzo della *zero-order correlation* tra i due tipi di attivazione; questo è spiegato con il fatto che la semplice correlazione non è in grado di discriminare chiaramente tra i due tipi di *arousal* perché questi modelli sono consistenti con lo stesso schema con un'analisi di correlazione di questo tipo: correlazione positiva tra *valence* ed *energetic arousal*, correlazione negativa tra *valence* e *tense arousal*, e indipendenza tra i due tipi di *arousal*. In contrasto, le correlazioni tra i residui dei due tipi di attivazione dopo aver sottratto la componente di varianza per la

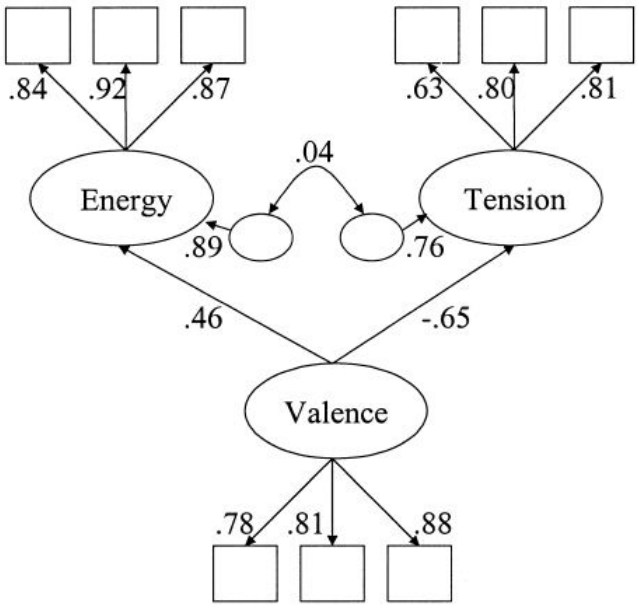


Figura 2.3: Correlazione tra i residui dell'*energetic arousal* e del *tense arousal*, senza la varianza condivisa in *valence*. (Schimmack e Reinsenzein, 2002)

valenza porta ad avere un maggiore dettaglio per quanto riguarda l'indipendenza o meno dei due fattori in considerazione, come spiegato sopra.

Il test proposto da Schimmack è stato pensato per mettere alla prova la teoria dello spazio *valence-activity*. Come prima analisi è stata rifatto uno studio di correlazione tra i tre fattori considerati, e ciò ha confermato i risultati già presenti in letteratura, con una correlazione positiva tra *energetic arousal* e *valence* (r = .46), indice di correlazione negativo tra *tense arousal* e *valence* (r = -.65), e debole correlazione negativa tra i due differenti *arousal* (r = -.28). Il contributo importante offerto da Schimmack sta ne fatto della valutazione sperimentale della correlazione tra i residui dei due tipi di attivazione, che come si può vedere in Figura 2.3 hanno una bassissima correlazione positiva (r = .04).

L'ipotesi di avere un modello bidimensionale basato su valenza e attivazione, per questo test, avrebbe dovuto comportare un indice di correlazione tendente a 1; i risultati, invece, confermano la teoria secondo cui ci sono due tipi indipendenti di *arousal*, ossia un modello bidimensionale per la sola componente di attivazione. Un punto su cui porre l'attenzione per affermare la bontà dei risultati è il fatto che la correlazione tra i residui e i loro fattori di attivazione è superiore al 50%, e questo indica che la valenza non comporta la maggior parte della varianza delle misure di attivazione, e quindi la correlazione tra i residui restituisce effettivamente una relazione tra componenti fondamentali della struttura dei due *arousal;* quindi la correlazione tendente a zero dei due residui non

può essere attribuita a una mancanza di varianza nei residui.

Questo studio supporta, ancora una volta, il fatto che i due tipi di attivazione si strutturino in maniera bidimensionale; non sostiene però che la dimensione relativa alla valenza sia composta da queste due componenti indipendenti; quindi come logica supposizione, lascia aperta la possibilità di una struttura a tre dimensioni non riducibile a sole due componenti base.

2.2.3 Sjoberg

Uno studio molto citato e confermato da successive analisi, è stato portato a termine da Sjöberg (1979). Questo ricercatore svedese, insieme ai suoi collaboratori, ha proposto un test composto da 86 aggettivi di tipo emotivo, a cui i partecipanti dovevano dare una valutazione su una scala simmetrica con due categorie di accettazione e rifiuto rispettivamente. Il test suddetto è stato sottoposto a un campione estremamente elevato di partecipanti volontari, in due test differenti, con il secondo che ha posto particolare attenzione ad avere eterogeneità in relazione all'età dei soggetti e alla loro professione.

Inizialmente un analisi esplorativa è stata eseguita attraverso metodi di analisi fattoriale, che permette di ottenere una riduzione della complessità del numero dei fattori che spiegano un dato fenomeno. Succesivamente è stata calcolata una correlazione multipla tra i fattori latenti trovati nel primo passo. Il problema di determinare quanti fattori siano adatti per la rotazione successiva è stato risolto attraverso un criterio di uscita che sfrutta il fatto di avere un gran numero di campioni, con grandi similitudini, e un gran rapporto tra il numero delle variabili e il numero dei fattori ipotizzati, com'è per questo caso. Il risultato dell'analisi fattoriale ha dato un risultato tra 6 e 7 fattori; sono poi stati considerati 6 fattori comuni perché indicano un ottimo assorbimento della varianza comune, con il 78% infatti della varianza stimata.

I sei fattori a cui entrambi gli studi hanno portato, sono qui riportati, con la distinzione che pur essendo tutti prettamente di connotazione bipolare, i primi 3 sono riconducibili a tre fattori di base per la struttura delle emozioni, mentre gli ultimi tre sono più associabili a caratteristiche di contorno:

- *Pleasantness*
- *Activation*
- *Tension*
- *Social Orientation*
- *Social interaction motive*
- *Control*

Questi ultimi 3 fattori latenti non sono stati considerati dagli autori come dimensioni basilari dello stato emotivo perché pur facendone parte in misura minore, sono definiti in termini di situazioni sociali in cui l'individuo viene a trovarsi;

molti altri stati di carattere soggettivo potrebbe altrimenti essere definiti, aspetti relativi prettamente a dimensioni che riflettono dimensioni base per quanto riguarda più l'aspetto di relazione interpersonale che emotivo. Gli autori quindi propendono per una struttura a due dimensioni, però quest'ultime misurabili secondo tre fattori principali.

2.2.4 Matthews

Lo studio fatto da Matthews e i suoi collaboratori (1990) considera tutta una serie di risultati presenti in letteratura, li analizza per quanto riguarda la metodologia usata, e propone la sua soluzione, che si presenta in linea con quanto affermato da Sjöberg (1979).

Il primo punto di analisi è il criterio utilizzato per determinare il numero dei fattori comuni o componenti principali dopo aver analizzato i dati. Tra i più importanti in letteratura, Russell utilizza il *K1 Criterion*, Watson e Tellegen, Thayer e Sjöberg usano lo *scree test*, o un criterio simile a quest'ultimo. Per primo Matthews sottolinea come il criterio K1 usato da Russell siano stato dimostrato essere inefficiente e con grossi risultati non corretti (Zwick e Velicer, 1986); invece, per quanto riguarda lo *scree test*, gli stessi Zwick e Velicer supportano empiricamente che questo criterio è corretto per quasi il 60% delle volte, con piccoli errori di sovrastima per una o due componenti risultanti.

Un secondo aspetto importante valutato in questo lavoro è la scelta del formato della risposta usato dai partecipanti, ossia il fatto di offrire ai soggetti un uguale numero di categorie di accettazione e negazione oppure in modo asimmetrico. L'autore su questo punto porta in esempio casi con tecniche simmetriche e non, e infine stabilisce per la sua prova l'uso di una scala di termini simmetrica, dovuta al fatto che una scala asimmetrica comporta poi in analisi una distribuzione asimmetrica dei risultati, con una certa dose di distorsione per quanto riguarda l'intercorrelazione tra le variabili in valutazione. Inoltre non utilizza la categoria definita come *cannot decide* in quanto comporta valutazioni non sempre semplici della stessa da parte dei soggetti (Cox e Mackay, 1985).

Matthews propone quindi un formato di risposta simmetrico, con l'uso in analisi del *Very Simple Structure Criterion* (VSS) invece del non perfetto *scree test* (Revelle e Rocklin, 1979).

La soluzione risultante dall'analisi fattoriale eseguita sulle basi del criterio VSS supporta la teoria di un modello tridimensionale dell'emotività (Figura 2.4). Le etichette date a ciascuno dei tre fattori sono rispettivamente *energetic arousal*, *tense arousal* e *hedonic tone*. La correlazione calcolata tra i due tipi di attivazione molto bassa (r = .03), mentre l'indice di correlazione è positivo tra *hedonic tone* e *energetic arousal* (r = .40) e negativo tra *hedonic tone* e *tense arousal* (r = -.30).

La conclusione, per quanto riguarda suddetti risultati, afferma che il modello tridimensionale meglio si adatta all'analisi della struttura dell'emotività, e afferma che risultati precedenti che sostengono un tipo di modello con solamente due fattori possono essere dovuti a una non corretta analisi fattoriale e a uno scorretto sistema di rilevamento dati; inoltre, nel caso del modello ricavato da

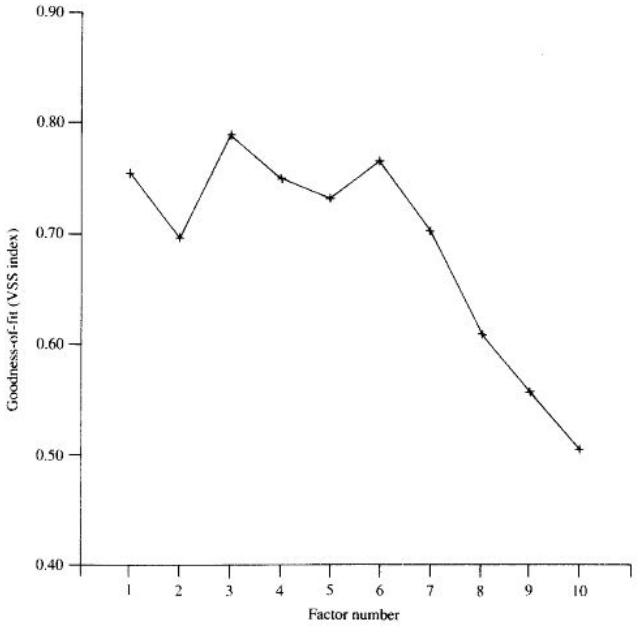


Figura 2.4: Indice *Goodness-of-fit* VSS in relazione al numero di fattori latenti estratti. (Matthews et al., 1990)

Thayer (1989), l'autore ha posto in studio la struttura particolare dell'*arousal*, non approfondendo la possibilità di ulteriori fattori. I dati sono stati confermati anche da successive analisi con campioni di popolazione differente, evitando perciò un'incoerente analisi dovuta a fattori di tipo demografico e di tratti personali diversi.

Capitolo 3

Emotional Computing

3.1 Intenzioni espressive nella musica

Negli ultimi decenni *human-computer interaction* (HCI) ha acquistato un ruolo di notevole interesse per quanto riguarda la ricerca e in ambito applicativo; la possibilità di poter comunicare contenuti espressivi da parte delle macchine, e in particolare computer, ha creato un intenso interesse attorno allo studio stesso della comunicazione tra umano e umano, per poter poi riprodurre tramite algoritmi in maniera sistematica.

I ricercatori partendo da modelli teorici, come visto nel capitolo precedente, hanno iniziato a sviluppare tutta una serie di algoritmi di sintesi e analisi dei contenuti espressivi (Volpe, 2003). In alcuni casi la ricerca è stata parte di settori già presenti nell'area della *computer science*, come a esempio il campo dell'Intelligenza Artificiale, mentre in altre situazioni una nuova area è stata appositamente sviluppata, che a seconda della zona di origine si differenzia in parte dalle altre, e in particolare sono:

- *Affecting Computing*, negli Stati Uniti;
- *KANSEI Information Processing*, in Giappone;
- *Expressive Information*, in Europa.

Vengono esposte di seguito le differenze che contraddistinguono queste diverse discipline, per poi proseguire con un'analisi più accurata dei contenuti del settore europeo e alcuni risultati di particolare importanza per gli scopi di questa dissertazione.

3.1.1 Affective Computing

L'approccio americano all'espressività nell'interazione computer-uomo prende forma attorno a un omonimo libro, scritto dalla docente dell'MIT Rosalind Picard (1997). Nel suo libro Picard definisce un livello nuovo di computazione,

ossia una relazione stretta tra ciò che viene percepito dalla macchina e una risposta adeguata all'utente che la utilizza, basata sul primo concetto. Per l'approccio statunitense, l'*Affecting Computing* mira a modellare e implementare tre concetti di base, i quali sono:

- il riconoscimento delle emozioni;
- la possibilità di esprimere emozioni;
- la sperimentazione di emozioni proprie da parte delle macchine.

L'idea principale che è alla base di questa scienza è che queste macchine siano *human-centered*, cioè riescano ad avere la possibilità di esplorare le reazione di un utente, di interpretarle, e di rispondere adeguatamente attraverso appositi segnali agli stimoli prodotti. Si parla di "stato emozionale" della macchina proprio per sottolineare il fatto che la macchina oltre che percepire e reagire di conseguenza, si trova di volta in volta in una situazione emozionale propria indipendente a ciò che si sta producendo all'esterno, situazione naturalmente causata da precedenti situazioni esterne nate dall'interazione con l'utente.

Passiamo ora alla descrizione sommaria dei tre concetti che stanno alla base dell'*Affective Computing.*

Il riconoscimento delle emozioni vede macchine che siano in grado di collezionare una serie di stimoli esterni, una vasta gamma di segnali, come possono essere espressioni vocali, espressioni facciali o particolari comportamenti legati al movimento del corpo; negli ultimi anni inoltre, molta ricerca si è basata sulla comprensione delle emozioni in base ai segnali fisiologici a cui ognuno è strettamente legato. Dopo una prima fase di raccolta dati devono essere implementati algoritmi di estrazione delle features e di classificazione di questi impulsi, per riuscire a riconoscere attraverso il ragionamento computazionale lo stato emotivo del soggetto. Tecniche di *machine learning* si adattano particolarmente bene per classificare lo stato emotivo di uno specifico utente, con conseguente reazione della macchina, anche basata sul proprio precedente stato emotivo.

Per quanto riguarda la possibilità di esprimere emozioni, la macchina dovrà essere in grado di utilizzare segnali di vario genere, da sintesi vocale o musicale a segnali visivi come animazioni o cambi di colore; questi segnali dovranno essere integrati in un'implementazione di un modello che riesca effettivamente a riprodurre una chiara manifestazione emotiva comprensibile da parte dell'utente. Di particolare importanza in questa idea è il *feedback* utente-macchina, che migliora l'utilizzo corretto da parte del soggetto; quest'ultimo potrebbe infatti percepire solamente confusione dall'interazione e non essere invece aiutato nel suo lavoro.

L'ultimo vede la sperimentazione da parte della macchina, in base a meccanismi e a ragionamenti propri, di provare uno stato emotivo personale, non dovuto all'interazione esterna. Questo punto è stato molto discusso in letteratura, e non è di facile valutazione. Per prima cosa si deve definire precisamente che cosa significa la presenza di uno stato emotivo interno alla macchina, e tutte le caratteristiche su cui si basa; poi si devono valutare i ragionamenti computazionali che permettono a un computer indipendentemente di elaborare una

situazione emotiva propria. Infine, ma non ultimo a livello di importanza, è l'utilità di una caratteristica tale da parte di un soggetto umano; il fatto di poter avere un'interazione emotiva con la macchina è qualcosa di indispensabile per migliorare l'interazione da parte dell'utente con la stessa, ma l'introduzione di automatismi propri potrebbe portare a un confronto non sempre utile e piacevole per il soggetto, che avrebbe un problema ulteriore non dover comprendere e relazionarsi quasi alla pari con la stessa macchina.

Molti progetti nel campo dell'*Affecting Computing* vedono scenari differenti di applicazione, che vanno dall'educazione, all'intrattenimento, alla detenzione di stati emotivi a scopi terapeutici. Per esempio la possibilità di implementare un tutor computerizzato che adatti la spiegazione in base alla reazione di chi lo sta ascoltando è di fondamentale importanza per la migliore interazione con la classe; in molte situazioni la determinazione di uno stato emotivo non è semplice, basti pensare al caso di soggetti autistici, e un'interazione macchina-utente che sia implementata per la comunicazione e l'insegnamento potrebbe essere di notevole aiuto per il paziente; per finire viene citato il caso più importante rispetto a possibilità di mercato e di profitto, che è il campo dell'intrattenimento, vede per esempio l'interazione tra utente e software di gioco, o qualsiasi altra piattaforma in grado di apportare benessere alla persona attraverso attività ludiche.

Per concludere, quindi, gli aspetti di maggior interesse riguardanti l*'Affecting Computing* sono la possibilità di creazione di classificatori di emozioni, di elaborazione dati e di riproduzione di ciò che viene definita una emozione da parte della macchina, ma che altro non è che una ragionata collezione di stimoli e segnali studiata per riprodurre uno stato riconoscibile dall'utente.

3.1.2 KANSEI

Alternativamente, e nello stesso periodo di sviluppo dell'*Affecting Computing,* in Giappone si è sviluppato l'approccio all'espressività da parte delle macchine definito con il nome di *KANSEI Information Processing.* Come l'approccio americano faceva riferimento a Picard e al suo gruppo di ricerca all'MIT, anche questa seconda visione qui presentata ha tra i maggiori apporti il lavoro di Hashimoto (1997), il quale struttura questo processo dell'informazione in tre fasi principali, che sono:

- elaborazione dell'informazione a livello fisico: in cui i segnali vengono catturati dal mondo esterno, i quali possono essere di vario tipo, da luminosi a sonori; il campo dell'elaborazione dei segnali è il più adatto per questa prima fase procedurale;

- elaborazione semantica dell'informazione: fase in cui avviene tutta l'elaborazione ai fini di ricavare regole valide in generale e conoscenza dai segnali identificati nella prima fase; questa fase ben si presta a essere sviluppata nell'area dell'Intelligenza Artificiale, che già di per sè copre questi aspetti;

- *KANSEI Information Processing* (KIP): il concetto che sta alla base di questo livello di elaborazione dei dati è strettamente legato alla cultura giapponese; KANSEI è una parola giapponese che non ha una diretta traduzione con nessuna delle parole del vocabolario occidentale, o comunque, ogni tentativo di traduzione si limita a catturare solo alcuni aspetti di un concetto molto più esteso. Il concetto di KANSEI è fortemente collegato al concetto di personalità e sensibilità, è un'abilità che permette a un soggetto umano di risolvere problemi ed elaborare informazioni in maniera rapida e personale; in ogni azione eseguita da un soggetto umano si può riscontrare la presenza del KANSEI, che sia un'azione di risoluzione di un problema, o semplicemente un modo di pensare personale. KANSEI è fortemente connesso col concetto di emozione, anche se non sono prettamente sinonimi; il primo infatti è più specifico per quanto riguarda l'analisi e la sintesi dell'informazione.

  Esempi di questo concetto sono i seguenti: un artista esprime il proprio KANSEI attraverso le proprie opere o performance, lascia traccia del suo KANSEI nei suoi prodotti, nei suoi messaggi; un attore o una ballerina interpretano il KANSEI richiesto per il particolare personaggio che stanno rappresentando. Se consideriamo poi questo concetto non come un contenuto da proporre, ma come un mezzo per interpretare, allora un esempio è la persona che sta seguendo uno spettacolo di un ballerino o di un attore e utilizza il proprio KANSEI per valutarlo, per estrarre il significato e la propria interpretazione sul pezzo in esame.

Il *KANSEI Information Processing* riguarda quindi processi di codifica e di decodifica; il KIP determina un modello nel quale i contenuti espressivi sono concepiti come informazione di alto livello, a livello simbolico, in un ambiente di interazione tra esseri umani; questi contenuti simbolici vengono quindi codificati in modo da trasportare questo messaggio ulteriore aggiuntivo alla pura informazione.

Nel momento in cui un utente umano manda un messaggio a un altro utente umano, implicitamente aggiunge informazione espressiva al suo interno; tale informazione, insieme al contenuto simbolico del messaggio stesso, viene codificata nel segnale fisico. Quando il ricevente riceve il segnale lo decodifica ed estrae sia il contenuto simbolico che l'informazione espressiva che vi è contenuta, dandone una sua interpretazione. Questo scambio di contenuti espressivi non sempre avviene in maniera volontaria, molto spesso sono dovuti a un comportamento inconscio del soggetto.

Comparando l'approccio americano con il KIP, si può notare una certa compatibilità, ossia tutti e tre i concetti illustrati precedentemente nel modello dell'*Affecting Computing* trovano riscontro anche nel *KANSEI Information Processing.* Il mittente del messaggio esprime, volontariamente o meno, un'informazione espressiva aggiuntiva che può essere associata alla fase di espressione di uno stato emozionale; il ricevente invece riconosce le emozioni espresse decodificando il segnale fisico; infine, il concetto di uno stato emozionale posseduto dalla macchina come è proposto nella letteratura americana può essere visto co-

me una relazione bidirezionale tra i due soggetti, entrambi con un proprio stato emozionale, il quale influenza ad alto livello l'informazione espressa nei messaggi successivi.

Un esempio di studio e applicazione svolta da Suzuki e Hashimoto è la possibilità di realizzare un agente in grado di gestire l'interazione tra esseri umani e robots, incorporando un modello computazionale di emozioni artificiali con caratteristiche di apprendimento e auto-adattamento (Suzuki et al., 1998). L'azione gestuale dell'utente umano è in grado di cambiare lo "stato emotivo" del robot, che si traduce in un cambio di stile per quanto riguarda il movimento ed il comportamento del robot, espresso anche da una serie di telecamere, luci, musica che il robot può comandare (Suzuki et al., 2002). Questo progetto è considerato dai creatori un ottimo aiuto per la stimolazione della creatività umana, e non presenta l'aspetto negativo di dover indossare particolari sensori per interagire con la macchina.

3.1.3 Expressive information processing

La ricerca che supporta il processamento dell'informazione espressiva in Europa si specializza in molti casi versi domini artistici e culturali. La possibilità di interazione tra discipline così diverse crea nuove opportunità di crescita per entrambe. Dal punto di vista dei settori scientifici e ingegneristici la possibilità di usare modelli teorici analizzati e studiati ampiamente in psicologia, nelle scienze sociali e in quelle umanistiche, aiuta il processo di sviluppo e implementazione di strumenti in grado di elaborare questo tipo di dati; mentre i settori più umanistici e psicologici possono avere grande supporto dal settore scientifico grazie a tutta una serie di innovazioni tecniche che semplificano la raccolta dei dati o creano possibilità di analisi di segnali che prima non era possibile studiare.

Le discipline artistiche spesso sono gli scenari di maggiore applicazione per il motivo che in questi campi l'analisi e la sintesi dell'emozione e dell'espressività è di centrale importanza. Gli studi di psicologia riguardanti questo aspetto, vedi capitolo precedente, si rivolgono alla conoscenza dei costrutti che intervengono nell'elaborazione dell'emozione da parte di un ascoltatore dopo aver ricevuto un messaggio, elaborata e percepita personalmente e indipendentemente.

La ricerca nell'ambito dell'elaborazione espressiva è ed è stata una delle attività principali del Centro di Sonologia Computazione (CSC) dell'Università di Padova, fin dagli anni settanta. Molti sono gli aspetti approfonditi attualmente dal *Sound and Music Computing group*, diretto dal prof. Giovanni De Poli:

- la creazione di un'interfaccia grafica che permetta di controllare, ad un livello di alta astrazione, i contenuti espressivi e l'interazione tra l'utente e l'oggetto audio considerato (Canazza et al., 2004). Per realizzare il passaggio tra differenti intenzioni espressive vengono usati due spazi di controllo astratti, il primo derivante da un'analisi percettiva dei contenuti audio - obiettivo di questo lavoro è la migliore comprensione di questo spazio - ed il secondo che lascia la possibilità agli autori di organizzare in modo personale le intenzioni espressive nello spazio di controllo.

- Lo studio delle analogie esistenti tra suono e movimento, derivante dalla consapevolezza che il movimento gestuale dell'esecutore influenza l'intenzione espressiva percepita dall'ascoltatore (Fenza et al., 2005; Camurri et al., 2005). Il mapping tra i movimenti del performer e l'interpretazione delle intenzioni espressive del brano ha portato alla definizione di due spazi espressivi tridimensionali, il primo ottenuto dalla teoria dello sforzo di Laban e Lawrence (1947), ed il secondo derivante dall'analisi multidimensionale di test percettivi.

- La creazione di un modello di sintesi del suono che sintetizzi un'esecuzione musicale trasformandola da una performance neutra, esecuzione della partitura senza intenzioni espressive o scelte stilistiche, ad una espressiva; questa trasformazione si basa sulla costruzione di un modello percettivo che descriva come gli ascoltatori organizzano le differenti espressioni musicali (Canazza et al., 2003a).

Un ulteriore esempio in ambito di ricerca è uno dei lavori di Roberto Bresin, ossia lo sviluppo di teorie, modelli e strumenti per la rappresentazione dei movimenti umani tramite suoni (Bresin e Friberg, 2010); questo lavoro è parte integrante di un settore di ricerca in crescita conosciuto come *data sonification*, *embodied music cognition* e tecnologia di mediazione.

"*Computational models for the discovery of the world's music*" (CompMusic), è un progetto di ricerca a livello europeo che si prefigge una descrizione ed una formalizzazione maggiore nei confronti della musica, per renderla più accessibile ad approcci di tipo computazionale e ridurre il divario tra la descrizione del segnale audio e i concetti a livello di semantica musicale; rappresenta uno dei progetti attuali più importanti di Xavier Serra, e collaboratori, da anni attivi nel settore del *music computing* (Ramirez et al., 2011).

Lo scopo principale dei settori scientifici resta quindi di sviluppare e interpretare dei modelli in grado di descrivere l'interazione del contenuto espressivo in un messaggio e la sua successiva comprensione. Tale modello può appoggiarsi a più sorgenti di segnale, per avere un'immagine più chiara della situazione e un insieme di informazioni più vasto, e quindi il sistema diventa multimodale; la comunicazione multimodale è alla base della ricerca nel campo dell'espressività, rappresenta un miglioramento rispetto a una sola forma di comunicazione analizzata, ed è la comunicazione che nella maggior parte dei casi avviene tra due esseri umani.

3.2 Sistemi espressivi e interfacce

In questi ultimi anni i progressi fatti nell'elaborazione dell'informazione ha portato ad avere anche un aumento notevole dell'informazione da comunicare e da gestire. Basti pensare alla grande quantità di dati che serve ora per salvare un file video in alta definizione rispetto a un file video di qualche anno fa, o la quantità di byte riservati per una foto digitale. Questa evoluzione si è verificata anche nella comunicazione umana, in riferimento a una certa maniera di

atteggiarsi, di esprimersi, di utilizzare tutta una serie di modalità connesse a essa; il maggior sforzo richiesto a livello cognitivo per gestire tutto ciò invece non è aumentato per quanto riguarda il cosiddetto *canale affettivo*, con il quale intendiamo il mezzo di comunicazione dell'informazione affettiva che viene trasmessa in aggiunta all'informazione non affettiva; difatti l'informazione affettiva è prodotta nel maggior numero dei casi in maniera non cosciente, e può essere percepita in parallelo al messaggio, non richiedendo incremento del carico di lavoro.

A seconda del canale usato per comunicare, quindi, vi è una possibilità minore o maggiore di riuscire a comunicare anche una componente espressiva implicita. Uno degli obiettivi principali della ricerca vede in particolare la possibilità di riconoscere l'espressività prodotta, codificarla in qualche tipo di forma standardizzata, di trasmetterla attraverso il canale in questione, e infine di ricodificarla al ricevente e interpretarla nella corretta maniera.

I criteri generali per una corretta costruzione di una modello o interfaccia che siano in grado di riconoscere le emozioni trasmesse sono (Volpe, 2003):

- gestione degli *input*: collezione di segnali a livello fisico che nel loro totale vanno a costruire l'intero messaggio di partenza, sia nella sua forma simbolica che espressiva;
- *pattern recognition*: estrazione di features a livello del segnale fisico e classificazione delle stesse;
- *reasoning*: predizione dello stato emozionale espresso, basato sulla classificazione delle features ma anche sul contesto, sulla situazione, sugli scopi e le preferenze personali, e su tutta una serie di altri aspetti associati alla conoscenza;
- *learning*: possibilità di discriminare il comportamento di un particolare soggetto rispetto a un altro, per il migliore adattamento riguardo al caso specifico; differenziazione dell'espressività;
- generazione di *output*: esternazione da parte della macchina del risultato di tutto il processo di riconoscimento dell'emozione.

3.2.1 Rappresentazione dell'informazione espressiva

L'espressività, comunicata in forma di informazione, può essere distinta secondo differenti livelli di astrazione. Il modello tradizionale vede la contrapposizione tra due livelli: un basso livello di rappresentazione, più vicino a come si presenta il segnale fisico, e un alto livello di astrazione nel quale l'informazione viene presentato a livello semantico (Figura 3.1, parte sinistra).

Secondo suddetto modello esiste una relazione di *mapping* diretta e bidirezionale tra i due livelli, ossia può presentarsi il caso di dover distinguere un particolare stato d'animo in base alla percezione di particolari features del segnale, ma può anche capire che per prima cosa avvenga una classificazione semantica

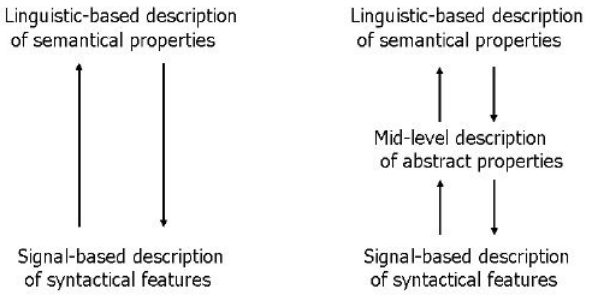


Figura 3.1: Livelli di rappresentazione dell'informazione espressiva. (Volpe, 2003)

dell'emozione e che si riporti poi a un basso livello di astrazione, con il riconoscimento dei segnali che l'hanno provocata. Il riconoscimento dell'emozione, quindi, non solamente tramite sviluppo *bottom-up* dal segnale al significato, ma anche *top-down* da un livello di alta rappresentazione simbolica che influenza quindi i segnali elaborati. Comunque, dal punto di vista dell'elaborazione dell'informazione, è conveniente definire un terzo livello di astrazione, intermedio agli altri due, il quale può essere diverso a seconda dei contesti (Figura 3.1,parte destra); questo modello intermedio sarà analizzato in dettaglio nel proseguo di questo capitolo, e servirà da punto di partenza per l'analisi compiuta e spiegata in questa tesi.

Interessante a questo punto è un esempio completo su come si sviluppi il riconoscimento di un'emozione quando siamo di fronte a un nostro interlocutore. Per prima cosa vengono valutati tutta una serie di segnali basilari, come alcuni movimenti del corpo, l'espressione del viso e degli occhi, la gestualità, il tono della voce e le parole usate per esprimere il concetto; quelli citati sono segnali che comportano informazione espressiva, e sono valutabili da un essere umano, altri invece, come la pressione sanguigna o il livello ormonale non sono direttamente visibili senza l'ausilio di apparecchiature apposite. Come seconda azione vengono poi combinati tra loro questi segnali in modo da poter essere associabili più facilmente a un concetto simbolico di più alta astrazione; una combinazione particolare di gestualità delle mani associata a un cero tono di voce può essere interpretato con uno stato di rabbia. Questo livello intermedio di rappresentazione dei *pattern* è utile per la migliore decisione finale riguardo a quale emozione sia rappresentata.

In Figura 3.2 vengono descritte le fasi essenziali in un progetto di modello per il riconoscimento dell'informazione espressiva da un segnale. A livello fisico, come già descritto, avverrà una estrazione delle features necessarie, che poi vengono mappate su pattern riconoscibili a livello intermedio; infine, dati questi pattern, verranno classificati, attraverso un'elaborazione basata su modelli teorici ben definiti, e il risultato sarà il contenuto espressivo del messaggio, descritto in termini basati sul linguaggio che descrivono le proprietà semantiche

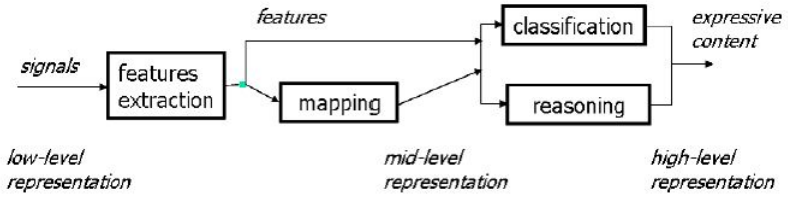


Figura 3.2: Modello per il riconoscimento dell'informazione espressiva. (Volpe, 2003)

dell'emozione.

3.2.2 Intenzione espressiva nella performance musicale

In ambito musicale, la comunicazione tra la trasmissione e la ricezione è costituita da numerosi elementi, in relazione anche al repertorio musicale. Nella musica tonale occidentale ci sono inclusi il compositore, la partitura, l'esecutore, il segnale acustico e per finire l'ascoltatore finale; se consideriamo invece la musica elettronica i componenti cambiano, il compositore agisce direttamente sul segnale e quindi la figura dell'esecutore non è più necessaria; nella musica di improvvisazione, ancora, il ruolo del compositore si fonde con la sua interpretazione esecutiva, relegando a un ruolo secondario la battitura, se presente.

Lo scopo di questa tesi è scoprire ulteriori informazioni per quanto riguarda la comunicazione tra le intenzioni dell'esecutore e l'esperienza provata dall'ascoltatore, ossia la percezione di quest'ultimo soprattutto per quanto concerne gli aspetti della comunicazione di contenuti espressivi, sia sottoforma di sensazioni che di emozioni.

La musica può esprimere emozioni in molti modi differenti: le emozioni possono essere collegate a situazioni che si stanno vivendo, possono essere legate a scostamenti dall'attesa che ha un ascoltatore, o ancora riflettere lo stato emozionale dell'esecutore e dell'ascoltatore. Un evento musicale si compone di tutte le caratteristiche elencate, è un evento non semplice da studiare e comprendere in tutte le sue sfumature. Molti studi sono presenti in letteratura per quanto riguarda lo studio della percezione delle emozioni nella musica; soprattutto sono ricerche svolte con estratti musicali proveniente dal repertorio occidentale classico, e il punto confermato da tutti e di partenza per l'analisi è il fatto che vi sia una stretta relazione tra un estratto musicale, seppur di breve durata, e un ampia gamma di emozioni che può trasmettere.

3.2.2.1 Influenza della struttura musicale sull'espressione emozionale

La ricerca più empirica si è basata sull'espressione musicale in modo da trovare quali emozioni siano maggiormente esprimibili nell'ambito musicale, ma anche

ha posto in evidenza quali siano i fattori che contribuiscono alla percezione dell'espressione musicale; i fattori citati sono relativi alla struttura della composizione musicale, associati quindi alla notazione musicale, come il tempo, la tonalità, la melodia, l'armonia, e varie proprietà di tipo formale. Seppur sia di concezione popolare il fatto che un compositore esprime le proprie emozioni attraverso la propria composizione, una visione più plausibile è che il compositore cerchi di usare vari fattori strutturali per raggiungere l'espressività cercata, in maniera differenziata tra lavori diversi, con o senza connessione con le emozioni provate in quel preciso momento.

Solitamente l'ascoltatore è posto in situazioni di poter giudicare non la diretta composizione, ma una sua interpretazione derivante dalla performance dell'esecutore, il quale modifica ulteriormente i fattori di base della battitura per dare una propria espressività al brano che sta suonando. L'esecuzione, quindi, comporta varie modifiche alla struttura, per esempio variazioni del tempo, della dinamica, dell'articolazione, dell'intonazione, ecc.

L'espressività percepita dall'ascoltatore, quindi, dipende sia da fattori dovuti alla struttura compositiva, sia da fattori dovuti al tipo di esecuzione. La maggior parte delle esecuzioni comportano delle intenzioni espressive derivate dalla volontà dell'esecutore, quindi l'interpretazione di queste intenzioni viene svolta dal lato dell'ascoltatore; gli approcci per la definizione di una struttura in grado di interpretare le intenzioni espressive sono stati presentati nel capitolo precedente, con la spiegazione di vari modelli teorici risultanti. Nella prossima parte di questo capitolo verrà presentato un modello di astrazione intermedio (Figura 3.1) in grado di interpretare l'espressività comunicata nella musica, in particolare basandosi su aggettivi di tipo sensoriale.

3.2.3 Kinestetic space

In letteratura esistono molti studi finalizzati alla definizione di spazi con basse dimensionalità che siano in grado di rappresentare al meglio l'intenzione espressiva dei vari brani musicali. Un esempio di mapping per l'analisi di performance musicali è il modello proposto da Langner e Goebl (2003), il "PerformanceWorm", nel quale la dimensionalità sfruttata è di due assi, il primo relativo al tempo e il secondo alla dinamica, la cosiddetta *loudness*; questa mappa mostra in tempo reale l'evoluzione di questi parametri sonologici di riferimento durante l'esecuzione del brano, generando un output continuo su una mappa bidimensionale che poi sarà sfruttato per analizzare l'andamento del brano.

La maggior parte delle ricerche che coinvolgono la definizione di spazi a bassa dimensionalità per l'espressività musicale sfruttano l'analisi delle componenti sonologiche degli estratti. Il modello qui presentato e proposto da Canazza (2003b) vede la definizione di uno spazio dimensionale in relazione ad aggettivi di tipo sensoriale.

La metodologia utilizzata è riassunta in Figura 3.3. Sono state dapprima valutate, attraverso esperimenti di tipo percettivo, una serie di esecuzioni basate su intenzioni espressive diverse. La fase di comprensione consiste nel derivare una struttura a bassa dimensionalità attraverso metodi di analisi multivariata dei

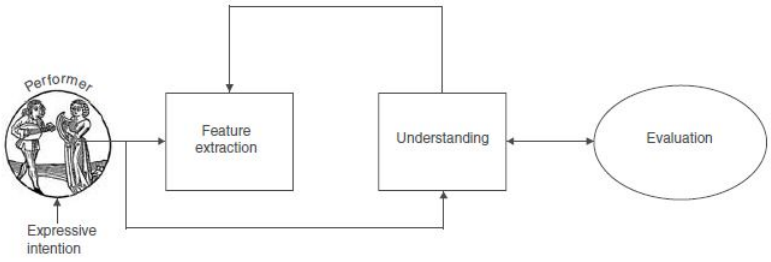


Figura 3.3: Metodo per l'analisi delle intenzioni espressive nel dominio sensoriale. (Canazza et al., 2003b)

dati. Infine ogni performance è stata analizzata per l'estrazione delle features acustiche che l'esecutore ha modificato per comunicare le differenti espressioni; queste features sono state infine utilizzate per la comprensione del modello dimensionale ricavato.

Il passo che comporta l'estrazione delle componenti del segnale può essere usato, inoltre, per determinare le features rilevanti da inserire in ingresso a un sintetizzatore espressivo (Canazza et al., 2000). Sfruttando il fatto che il sintetizzatore espressivo opera sulle features e non sul segnale, è possibile per esempio comprendere il dominio dell'esecuzione musicale, e applicare le operazioni a processi particolari per quanto riguarda il campo dell'elaborazione del segnale visivo, come nel caso di un movimento di un ballerino.

Nel lavoro proposto da Canazza, inoltre, è importante sottolineare l'importanza di aver fatto interpretare a degli esecutori professionisti lo stesso pezzo con intenzioni espressive differenti. Non sempre l'interpretazione espressiva di un brano viene accentuata con connotazioni emotive così forti. Ci sono però alcuni generi musicali che sfruttano questa possibilità di interpretazione per aumentare il coinvolgimento dell'ascoltatore, come succede nella musica jazz; un esempio di questa libera interpretazione è stato Miles Davis che a seconda della reazione del pubblico riusciva a modificare lo stile espressivo delle proprie esecuzioni. Ritornando all'esperimento, è stato di fondamentale importanza poter avere intenzioni espressive diverse in relazione allo stesso brano per poter poi far valutare ai partecipanti la corretta definizione delle stesse.

Gli estratti selezionati per l'esperimento sono stati eseguiti da esecutori professionisti secondo intenzioni espressive diverse, da un punto di vista sensoriale, seguendo gli aggettivi: *light, heavy, soft, hard, bright* e *dark*; inoltre è stata suonata un'esecuzione neutra per ciascun brano. Da notare che ogni aggettivo considerato ha il suo opposto, per esempio leggero e pesante, assicurando in questo modo il corretto contrasto tra le espressioni opposte.

Ai partecipanti è stata chiesta una valutazione quantitativa, in rispetto a ogni aggettivo, per ogni esecuzione musicale presentata. Una doppia analisi fattoriale è stata poi applicata a questa analisi percettiva usando come variabili le esecuzioni e gli aggettivi valutati. Un'analisi multidimensionale di tipo MDS

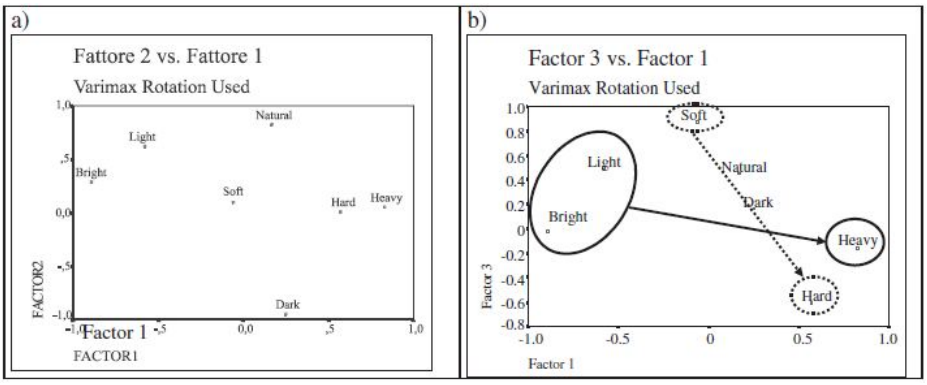


Figura 3.4: Risultato dell'analisi fattoriale, esperimento 1. (Canazza et al., 2003b)

è stata utilizzata per scoprire in quale maniera le varie esecuzioni siano state distinguibili tra loro, e in che modo possano mantenere le distanza in uno spazio a dimensionalità ridotta rispetto alle variabili di partenza. Infine una cluster analysis è stata necessaria per definire se ci siano situazioni nelle quali alcuni aggettivi siano vicini nello spazio dimensionale trovato, e quindi siano valutati pittosto similmente dai soggetti partecipanti all'esperimento.

I risultati ottenuti hanno proposto una dimensionalità ottimale per quanto riguarda le tre dimensioni (Figura 3.4); lo spazio così ottenuto rappresenta un modello per l'espressività, nel quale i partecipanti hanno organizzato le loro percezioni dei brani. L'analisi acustica delle esecuzioni (Canazza et al., 1997) mostra come il Fattore 1 sia strettamente correlato con il Tempo, ed è interpretato come fattore cinematico; il Fattore 3 invece è in relazione al tempo di attacco, al Legato/Staccato, all'intensità, ed è quindi interpretato come fattore energetico. Da questo la definizione dello spazio *Kinematics-Energy*.

Un ulteriore esperimento è stato proposto per convalidare i risultati ottenuti, sfruttando una lista di aggettivi molto più ampia, con l'intento di osservare e analizzare la percezione da parte dei soggetti anche dopo aver inserito tutta una serie di aggettivi intermedi, sempre nell'ambito sensoriale. In questo caso i due fattori dominanti hanno comportato una varianza cumulativa pari al 75.2%; inoltre, dall'analisi fattoriale eseguita è risultato chiaramente che i soggetti sono riusciti anche in questo caso a identificare le intenzioni espressive comunicate. Una seconda interpretazione dello spazio trovato è quindi la relazione dell'asse delle ascisse "*bright-light* vs *heavy*" come componente cinematica, mentre l'asse delle ordinate "*soft* vs *hard*" come componente di energia (Figura 3.5).

Lo spazio bidimensionale così proposto è stato usato come interfaccia per l'interazione uomo-macchina nella sintesi di esecuzioni musicali con differenti intenzioni espressive. Un modello per la mappatura dei punti nello spazio *Energy-Kinematics* è stato proposto, usando i risultati che relazionano alcune features ai fattori rilevati, come per esempio il Tempo, Legato e l'Intensità.

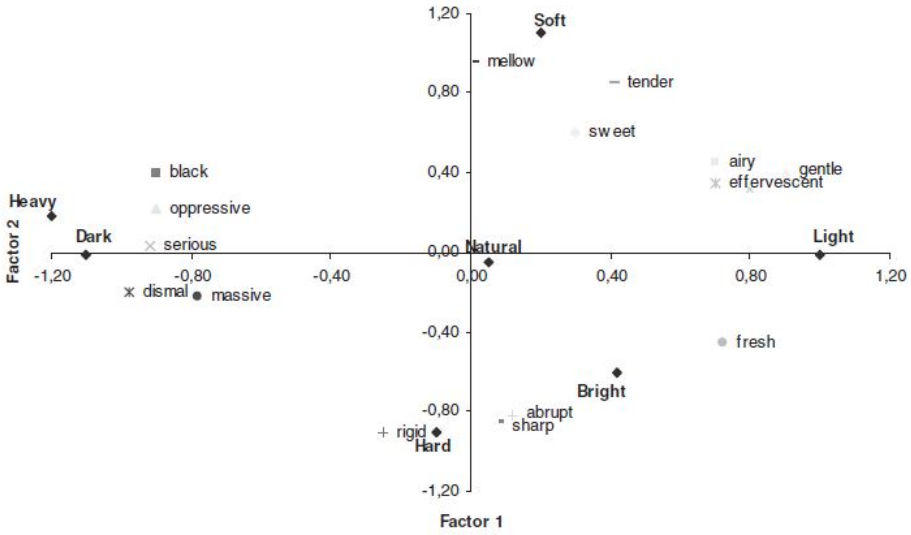


Figura 3.5: Risultato dell'analisi fattoriale, esperimento 2. (Canazza et al., 2003b)

Quindi, per presentare i contenuti espressivi, il modello usa una serie di controlli, che sono stati dimostrati essere più rappresentativi e indipendenti dallo strumento e dall'estratto musicale. Il sistema calcola questi parametri e determina la deviazione che deve essere applicata alla partitura per riprodurre una particolare intenzione espressiva. Il sistema è usato per via direttiva, ossia una particolare intenzione espressiva viene selezionata sullo spazio bidimensionale e il sistema suona la partitura con le intenzioni date dall'utente.

Capitolo 4

Esperimenti percettivi: stato dell'arte

In questo capitolo vengono presentati i più rilevanti risultati presenti in letteratura. Nella prima parte verrà esposto il lavoro di Emmanuel Bigand (2005), professore di psicologia cognitiva dell'università di Borgogna, Francia; la ricerca di Bigand si è proposta di studiare le caratteristiche simili che comportano intenzioni espressive percepite e associate tra loro dai partecipanti, e i risultati trovati sono stati di fondamentale importanza per l'analisi fatta in questa tesi, che può essere vista come una naturale continuazione dell'indagine iniziata da Bigand stesso; l'approccio seguito per questo lavoro è di tipo *score-dependent*, ossia un'analisi delle informazioni espressive della musica basata sulla partitura.

Nella seconda parte del capitolo sarà considerato, invece, un lavoro di ricerca proposto da Mion, De Poli e Rapanà (2010), nel quale viene studiata una possibile organizzazione comune per quanto riguarda le intenzioni espressive in campo affettivo e in campo sensoriale; questi ultimi modelli percettivi, infatti, vengono solitamente studiati e analizzati separatamente in tutti i loro aspetti, e quindi è di particolare importanza cercare una relazione che accomuni entrambi per comprendere in maniera ancora più completa come un soggetto organizzi la propria percezione relativamente alle esecuzioni musicali. L'approccio seguito, in questo caso, è di tipo *score-independent*, ossia una elaborazione dell'espressività della musica legata alla performance musicale, che si differenzia dall'approccio seguito da Bigand e introdotto precedentemente.

4.1 Reazione emotiva alla musica

Nei capitoli precedenti si è posto l'accento sul dibattito presente in letteratura per quanto riguarda la natura delle emozioni indotte dalla musica. Molti studi hanno utilizzato per i loro esperimenti di tipo percettivo una valutazione sulla base di termini linguistici di uso comune. L'uso di questa serie di termini, molto usati nella lingua sia parlata sia scritta, ma poco precisi, è potenzialmente pro-

blematico per la valutazione del soggetto partecipante nei riguardi di un brano che sta ascoltando; potrebbe crearsi un problema di semplificazione dell'esperienza emotiva provata da parte del soggetto, con difficoltà a differenziare in modo adeguato termini vicini tra loro come classificazione (Scherer, 1994).

Sulla base di questo pensiero, Bigand ha proposto una modalità procedurale che non faccia affidamento su classificazione su base semantica (Bigand et al., 2005). In particolare, quello che è stato richiesto ai partecipanti, è stata la totale attenzione all'esperienza emotiva provata durante l'ascolto, e in secondo luogo un raggruppamento degli estratti ascoltati in base alle differenti emozioni provate. E' importante sottolineare che le istruzioni date ai partecipanti miravano a una attenzione nel riconoscimento dell'emozione provata nell'ascolto del brano piuttosto che nell'individuazione dell'informazione emotiva che in esso è stata codificata.

4.1.1 Emozioni indotte ed emozioni percepite

La differenza tra emozione indotta e percepita è un argomento di dibattito in ambito di ricerca (Gabrielsson, 2001); la semplice richiesta ai partecipanti di una maggiore attenzione alle emozioni provate non garantisce che ciascun soggetto riesca a valutare effettivamente lo stato emotivo provato nell'ascolto invece dell'intenzione espressiva con cui l'esecutore lo ha suonato. Non è d'altronde di facile distinzione se questi due aspetti siano effettivamente separabili o se siano due facce della stessa medaglia.

Un primo esperimento proposto da Bigand è stato messo in atto per comprendere in maniera migliore la differenza tra emozione percepita e indotta. A un gruppo di partecipanti è stato richiesto di valutare una serie di brani sulla base di una scala semantica bipolare; nella prima fase l'attenzione è stata posta nei confronti delle emozioni provate dai soggetti, ed è stata raccolta una autovalutazione che rispecchiasse questo; nella seconda fase invece, sempre con gli stessi soggetti, l'interesse ha riguardato le emozioni che dovrebbero essere percepite durante l'ascolto, quindi una valutazione meno personale e più razionale. E' stata riscontrata una correlazione elevata nelle due valutazioni, con $r(34)=.93$, $p<.001$[1]. L'interpretazione data dall'autore a questo esperimento ha confermato l'idea che i partecipanti non valutino in maniera differente i due aspetti, ma piuttosto ognuno di loro dia una valutazione che, a seconda dell'abilità di ciascuno, è più rivolta a una sensazione provata personalmente oppure a una valutazione più razionale dell'intenzione espressiva, o per finire, una combinazione di questi due fenomeni.

[1] In statistica inferenziale il valore p (o p-value) di un test di verifica d'ipotesi indica la probabilità di ottenere un risultato pari o più estremo di quello osservato, supposta vera l'ipotesi nulla (l'ipotesi che si vuole verificare nel test, in contrapposizione all'ipotesi alternativa). Talvolta viene anche chiamato livello di significatività osservato.

4.1.2 Struttura delle emozioni indotte dalla musica

La prima parte del lavoro svolto da Bigand e dai suoi collaboratori è rivolta alla comprensione della struttura dimensionale che è alla base della risposta emotiva provocata nei soggetti partecipanti in relazione all'ascolto di una serie di estratti musicali del repertorio occidentale classico. A un gruppo di 19 partecipanti, tra cui 10 con esperienza musicale e 9 senza studi in questo ambito, sono stati sottoposti a valutazione 27 estratti musicali di musica strumentale, selezionati appositamente da un gruppo di musicologi e psicologi per rispecchiare una ampia varietà di emozioni, e per essere rappresentativi dei diversi stili della musica classica occidentale: barocco, classico, romantico e moderno. È stata posta particolare attenzione in fase di selezione per avere una varietà di estratti relativi a diversi gruppi strumentali: solo, musica da camera e orchestrale.

La particolare cura nella selezione dei brani è importante per neutralizzare ogni possibile effetto di confusione tra la struttura insita nel brano e l'effettiva intenzione espressiva prodotta; la ricerca della similitudine per i partecipanti non deve essere resa troppo difficoltosa per il fatto stesso di proporre gruppi di brani troppo simili strutturalmente tra loro. Inoltre, per evitare situazioni in cui il giudizio possa essere influenzato da una conoscenza troppo approfondita del brano, sono stati selezionati estratti poco famosi, aspetto importante soprattutto per coloro che appartengono alla categoria dei musicisti.

Gli estratti così selezionati hanno una durata media di 30 s e corrispondono all'inizio di un *movimento musicale* oppure all'inizio di un tema musicale ben definito.

L'esperimento si è svolto in due fasi, una prima fase valutativa in cui ai partecipanti è stato richiesto di raggruppare i brani secondo i criteri esposti precedentemente, e una seconda fase con la stessa procedura della prima, eseguita a distanza di 2 settimane dalla prima prova.

4.1.2.1 Risultati

È stata riscontrata una media di 8 gruppi formati per entrambe le fasi dell'esperimento, e senza differenze sostanziali tra la categoria dei musicisti e quella dei non musicisti. La percentuale dei brani posti in gruppi differenti nelle due sessioni è stata non troppo elevata, con un valore maggiore per i non musicisti (15%) rispetto alle scelte dei musicisti (9%); questo può essere imputato a una maggiore attitudine dei musicisti per quanto riguarda l'ascolto musicale stesso, una maggiore preparazione e quindi una interpretazione più decisa dei brani che resta anche nel tempo; per contro i non musicisti possono essere stati meno determinati nel raggruppamento di particolari brani con intenzioni espressive poco chiare.

È stata poi calcolata la matrice di dissimilarità per entrambe le fasi. Le matrici ottenute per le sessioni 1 e 2 hanno avuto un alto indice di correlazione, $r(349)=.87$, $p<.001$ per i musicisti e $r(349)=.78$, $p<.001$ per i non musicisti. Data questa alta correlazione nelle valutazioni per le due sessioni e per le due

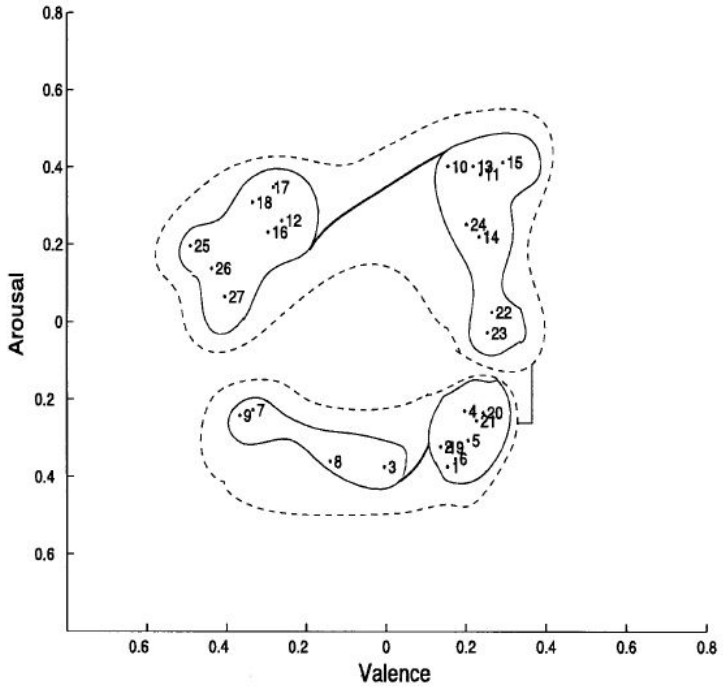


Figura 4.1: Rappresentazione risultante dall'analisi MDS e *cluster analysis.* (Bigand et al., 2005)

categorie di partecipanti, le quattro matrici sono state fuse tra loro, ed è stata fatta un'analisi multidimensionale MDS.

La riduzione dimensionale effettuata dall'analisi MDS ha portato ad avere come soluzione migliore uno spazio a tre dimensioni; aggiungendo la terza dimensione al modello l'*explained variance* aumenta dal 73% al 83%. La rappresentazione dei brani lungo le due dimensioni principali del modello è riportato in Figura 4.1. L'asse verticale oppone estratti musicali che variano rispetto al livello di *arousal*, con un livello che decresce dall'alto in basso, mentre l'asse orizzontale oppone brani che differiscono per valenza emotiva, maggiore a destra e minore a sinistra nel grafico. Il terzo asse, non visualizzato in figura, oppone, secondo la valutazione dell'autore, brani con contorni melodici ampi a brani che procedono armonicamente o con arpeggi spezzati.

Per valutare la bontà dei risultati ricavati dai test è stata eseguita una *bootstrap analysis*; 30 valutazioni scelte casualmente sono state analizzate come spiegato in precedenza, e questo per 200 iterazioni; ogni spazio ricavato per ogni iterazione è stato poi ruotato verso la stessa rappresentazione degli assi. La posizione di ogni brano in questa analisi ripetuta propone una nuvola di punti più o meno ampia a seconda della coerenza di valutazione tra i partecipanti, una maggiore concentrazione rispecchia una valutazione concordante tra i soggetti.

I risultati confermano la validità dei dati raccolti.

Per finire è stata proposta un'analisi di raggruppamento, o *cluster analysis*, per lo spazio risultante dall'analisi multidimensionale. Sono stati definiti quattro gruppi principali (Figura 4.1), che variano lungo differenti valori di *arousal* e *valence*: un gruppo associato alla sensazione emotiva di felicità, definito da un alto livello di *arousal* e valenza positiva; un secondo gruppo interpretato come serenità, con bassa dinamica ma ancora con valenza positiva; un terzo insieme denominato con i termini paura e rabbia, caratterizzato da alta dinamica e valenza negativa; infine un ultimo gruppo relativo allo stato d'animo di tristezza, con valori bassi in entrambe le dimensioni principali.

4.1.2.2 Discussione

Questa prima prova effettuata da Bigand ribadisce un risultato proposto molte volte in letteratura, ossia il fatto che le due dimensioni principali che strutturano la risposta emozionale delle persone all'ascolto di brani musicali sono: *arousal* ed *emotional valence*. L'autore indica inoltre come la dimensione della *valence* non sia correlata a giudizi di piacere; questo risultato suggerisce come, per esempio, la musica triste non sempre sia associata a emozioni di dispiacere, e viceversa.

Un aspetto importante per gli scopi di questa tesi è il fatto che questo esperimento ha messo in luce una terza dimensione che contribuisce alla percezione dell'intenzione espressiva della musica. Una possibile interpretazione dell'autore rispetto alle caratteristiche di questo asse, è che questo esprima l'influenza della gestualità corporea evocata dalla percezione delle emozioni che lo stimolo induce. Questa relazione tra musica e movimento è ben definita, e molti autori hanno enfatizzato che l'emotività musicale risulti in parte relazionata a modelli fisici di postura e gestualità (Damasio, 2003). Un ampio arco melodico evoca una gestualità lenta, ampia e continua, mentre un arpeggio spezzato può evocare movimenti discontinui. L'autore non approfondisce lo studio di questa terza dimensione per difficoltà di spiegazione utilizzando termini linguistici, e lascia l'analisi aperta a ulteriori lavori.

Un importante risultato di questo esperimento, confermato dall'analisi proposta in questa tesi, è la minima differenza che comporta l'esperienza musicale e lo studio di questa disciplina per quanto riguarda la percezione che un soggetto può avere nell'ascolto di un particolare estratto; l'analisi separata delle due categorie di partecipanti ha evidenziato infatti un alto indice di correlazione tra i risultati. Questo aspetto mette in luce come l'organizzazione percettiva musicale di un essere umano non sia particolarmente influenzata da una maggiore o minore conoscenza musicale, in particolare in questo caso relativo al repertorio occidentale classico.

4.1.3 Reazione emotiva - processi cognitivi

Bigand (2005) effettua inoltre altri due esperimenti per verificare se la durata di un estratto musicale ne influenzi la valutazione da parte del soggetto. In

particolare la prova eseguita ha visto la stessa serie di brani considerata nell'esperimento spiegato nella sezione precedente, proposta a 40 partecipanti, metà musicisti e metà senza particolari conoscenze in merito; quello che è cambiato è stata la durata degli estratti, ora ristretta a solo un secondo di esecuzione.

Il primo di questa coppia di esperimenti si è svolto con le modalità viste in precedenza, due sessioni a intervallo di due settimana una dall'altra, raggruppamento dei brani a seconda dell'emotività provata nell'ascolto. L'analisi MDS dei risultati ha riproposto una soluzione ottimale con tre dimensioni, con una *explained variance* maggiore per i non musicisti (77%) rispetto ai musicisti (68%). La stessa interpretazione della terza dimensione data per il precedente esperimento è riportata anche per quest'ultimo.

Le differenze maggiori rispetto ai risultati di Figura 4.1 si ricavano dall'analisi di raggruppamento successiva all'MDS. Per quanto riguarda la categoria dei musicisti, sono stati identificati tre gruppi principali; la differenza sta nel fatto che i musicisti incontrano alcune difficoltà nel percepire una differente componente di valenza emotiva quando i brani presentano un basso *arousal*; difatti, dove in Figura 4.1 si riconoscevano due gruppi distinti con basso *arousal*, nell'esperimento con brani da 1 secondo i musicisti non differenziano i due insiemi. La condizioni riscontrata per quanto riguarda i non musicisti è simile, comporta però il fatto di una difficile comprensione, da parte di questi, nei confronti di estratti con differente valenza emozionale; infatti non vi è una particolare differenziazione lungo l'asse della valenza, ma solamente per quanto riguarda differenti livelli di *arousal*.

Vista la presunta difficoltà dei partecipanti nel percepire l'emozione musicale trasmessa con estratti di un solo secondo e di riuscire anche a raggrupparli per similitudine, Bigand propone un ultimo esperimento nel quale propone tutte le possibili coppie di estratti, e i soggetti ne devono dare una valutazione rispetto a una scala bipolare che varia da "differenti" a "simili". I risultati ricavati dall'analisi per questa prova confermano quanto trovato in quella precedentemente esposta; l'unica differenza degna di nota è una migliore sensibilità da parte dei partecipanti nel differenziare livelli diversi di valenza in condizioni di alto *arousal*. Questo viene interpretato come un miglioramento dato alla metodologia usata, che è leggermente più sensibile di un semplice raggruppamento, per quanto riguarda brani da 1 secondo di durata.

4.1.4 Discussione generale

Molti sono i risultati utili ricavati dal lavoro di Bigand, soprattutto per quanto riguarda il lavoro svolto in questa tesi, che sarà spiegato in dettaglio nei prossimi due capitoli. Riassumendo brevemente i punti principali, Bigand ha ricavato dalle analisi fatte i seguenti risultati:

- la risposta emotiva alla musica non è soggetta a differenze individuali, come può essere una diversa competenza in ambito musicale, o differenze anche tra soggetti appartenenti alla stessa categoria.

- brani musicali molto vicini per la collocazione nel piano bidimensionale dell'emozione differiscono in maniera sostanziosa per quanto riguarda la strumentazione e lo stile musicale; questa osservazione suggerisce che l'esperienza emotiva provocata dalla musica corrisponde a un livello più astratto della semplice categorizzazione musicale, e quindi non si sofferma su aspetti puramente strutturali.

- la numerosità dei gruppi creati dai partecipanti, e la consistenza degli stessi, indica che i brani non sono stati classificati relativamente a delle specifiche emozioni di base; piuttosto, un approccio dimensionale come quello utilizzato riesce a indicare in maniera più completa come si possa organizzare a livello mentale in un soggetto la valutazione di quale sia l'interpretazione espressiva.

- la presenza di una terza dimensione contribuisce a definire in maniera più completa il modello; questa dimensione viene interpretata, come visto in precedenza, come una mentalizzazione del movimento musicale, una gestualità collegata all'esecuzione del brano; questo punto, però, non è stato approfondito in questo lavoro.

- l'ascolto di solo 1 secondo di un brano è sufficiente per creare la corretta interpretazione dell'intenzione espressiva; nella maggior parte degli estratti un solo secondo di esecuzione corrisponde a un singolo accordo o a un singolo tono; le caratteristiche più importanti che aiutano questo tipo di classificazione sono fattori strutturali, come la modalità, la consonanza, l'orchestrazione, il ritmo, ecc.; quindi, anche la sola esecuzione di un secondo può imprimere nell'ascoltatore l'espressione musicale desiderata, e determina quindi una maggiore importanza nella performance dell'esecutore.

4.2 Organizzazione percettiva dei domini sensoriali e affettivi nelle intenzioni espressive della musica

L'organizzazione in spazi relativi a domini affettivi e sensoriali, nella valutazione dell'intenzione espressiva della musica, è sempre stata studiata tradizionalmente in modo separato. Nello studio di Mion e De Poli (2010) è stata analizzata in maniera più dettagliata la comunicazione di contenuti espressivi da parte dell'esecutore.

Mentre le scienze cognitive concentrano la loro attenzione soprattutto sui processi a livello mentale, l'approccio seguito da Mion per questo lavoro segue l'idea di Leman in cui il corpo è considerato mediatore tra la situazione e l'esperienza soggettiva. L'ipotesi enattiva ritiene che la conoscenza non derivi solamente da una percezione passiva, ma accumuli informazione anche attraverso la necessità di azione in un ambiente in cui si verifica un interazione di

tipo percezione-azione (Varela et al., 1991). Una maggiore attenzione all'azione può aiutare a comprendere in maniera più chiara gli aspetti soggettivi dell'esperienza musicale; negli ultimi tempi sono state fatte importanti scoperte in questa direzione, cioè nell'associazione tra il dominio musicale e il dominio fisico (Canazza et al., 2010).

Negli ultimi anni è cresciuto l'interesse nello studio della componente espressiva aggiunta da un esecutore a un brano suonato (De Poli, 2004). Alcuni aspetti sono fortemente relazionati allo stile musicale, come può essere il particolare stile dell'esecutore, all'influenza della diversa preparazione musicale, e alla situazione particolare in cui si svolge la performance, con un maggiore o minore coinvolgimento del pubblico.

Quando vengono analizzate differenti intenzioni espressive, un punto fondamentale è il fatto di considerare diverse performance dello stesso brano con differente espressività; per far ciò viene richiesto a un esecutore di suonare lo stesso brano musicale più volte, aggiungendo ogni volta l'intenzione espressiva che intende comunicare. Date queste diverse interpretazioni, il ricercatore ha i mezzi per analizzare le strategie utilizzate e per determinarne particolari caratteristiche (De Poli, 2004).

4.2.1 Features della musica espressiva non legate alla partitura

Uno studio del 2008 ha messo in luce le caratteristiche fondamentali con le quali vengono comunicate e riconosciute le diverse intenzioni espressive, non strettamente connesse alla partitura (Mion e De Poli, 2008); viene messo in risalto la componente espressiva dovuta alla performance musicale, approccio definito *score-independent*. Utilizzando tecniche di *machine learning* è stata trovata una serie di caratteristiche a livello acustico, particolarmente importanti per il riconoscimento delle differenti espressioni. Differenti strumenti sono stati presi in considerazione per le analisi. Gli esperimenti condotti da Mion hanno dimostrato una buona descrizione dei contenuti espressivi della musica utilizzando features acustiche e basate su un intero evento, tra le quali le più importanti si sono rilevate la *roughness*[2], NPS (*Notes Per Second*), PSL (*Peak Sound Level*) e il tempo di attacco (Figura 4.2).

Tali features generalizzano i risultati già presenti in letteratura nei quali parametri come tempo, intensità del suono e articolazione erano usati per la classificazione delle espressioni musicali nel repertorio classico occidentale.

Da sottolineare come sia stata riscontrata una notevole importanza per quanto riguarda la componente legata alla *roughness*. Inoltre per ogni strumento sono state aggiunte alcune features caratteristiche, per la fase di analisi, che hanno contribuito a un migliore riconoscimento dell'espressione; questo aumento di prestazione non è stato, però, dominante nella comprensione, e inoltre questa valutazione approfondita può essere eseguita solo dopo l'uso di tecniche

[2] Calcolata secondo il *Synchronization Index Model* di Leman. (Leman, 2000)

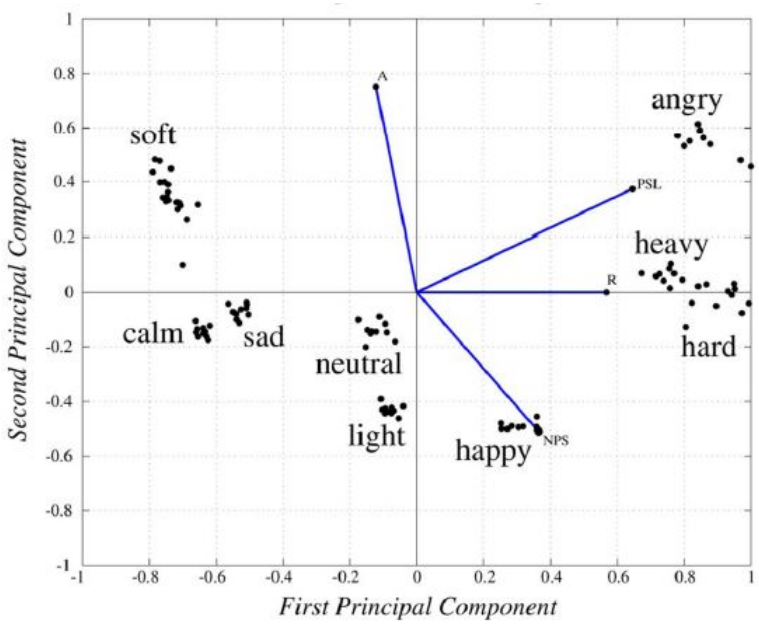


Figura 4.2: Analisi delle componenti principali per entrambi i domini affettivo e sensoriale, esecuzioni con chitarra. (Mion e De Poli, 2008)

di classificazione del timbro e quindi quando si sia già identificato il tipo di strumento.

La descrizione della comunicazione espressiva a un livello intermedio non si basa nè su aspetti derivanti dalla relazione con la partitura, nè dal linguaggio musicale; i risultati del lavoro di Mion sono quindi utili a questo scopo, ossia classificare le diverse espressioni musicali, e in applicazioni per la ricerca di contenuti espressivi su dati audio.

4.2.2 Organizzazione acustica e percettiva

A seguito dei risultati esposti in 4.2.1, considerando che le intenzioni espressive sono simili relativamente alle features calcolate per il loro riconoscimento, e ricordando che la fase di riconoscimento vede una forte relazione con la valutazione soggettiva, in (Mion et al., 2010) si analizza l'idea che le performance siano simili anche da un punto di vista percettivo.

L'attenzione viene posta nello studio di una possibile connessione tra organizzazione acustica e percettiva, per scoprire la relazione che esiste tra termini di dominio affettivo e termini di dominio sensoriale. Lo scopo dello studio qui esposto è riuscire a identificare un modello intermedio per il riconoscimento delle intenzioni espressive, che contenga features comuni ai due domini.

4.2.2.1 Sperimentazione e metodologia usata

Gli spazi considerati per l'analisi, che fanno riferimento alla letteratura, sono il modello bi-dimensionale *valence-arousal* di Russell (1980) e il modello nel dominio sensoriale definito da Canazza e analizzato nel capitolo precedente, il *kinematics-energy space* (Canazza et al., 2003b). Per analizzare entrambi questi domini sono stati considerati termini di significato opposto per definire le dimensioni di ogni spazio. Per lo spazio affettivo, la bipolarità indotta dalle dimensioni indipendenti viene rappresentata dalle contrapposizioni *happy-sad* e *angry-calm*, rispettivamente per distinguere livelli di valenza e di *arousal*. Per lo spazio sensoriale, le coppie *hard-soft* e *light-heavy* sono state scelte per rappresentare la variazione lungo le dimensioni dell'energia e della cinematica, rispettivamente. In questo modo ogni aggettivo ha il suo opposto, e comporta un'esecuzione contrastante dal punto di vista dell'espressività da parte dell'esecutore. E' stata presa in considerazione anche una performance con livello espressivo neutro.

I brani registrati hanno visto l'esecuzione di musicisti professionisti, con l'uso di tre strumenti diversi: violino, flauto e chitarra. In totale 324 brani sono stati registrati presso il Centro di Sonologia Computazionale dell'Università di Padova. Una prima analisi delle componenti principali e di raggruppamento delle features calcolate sui brani, e spiegate nella prossima sottosezione, ha messo in luce la definizione di tre cluster principali: (A) *hard/heavy/angry*, (B) *light/happy*, e (C) *sad/calm/soft* (Figure 4.3 e 4.4).

Per il primo esperimento è stato scelto un rappresentante per ogni cluster, usando tecniche di minimizzazione della distanza coseno; per i tre rappresentanti è stata considerata ognuna delle nove intenzioni espressive, per due strumenti diversi, per un totale di 36 estratti musicali. I partecipanti, composti da categorie di differente formazione musicale, hanno potuto classificare questi esempi di esecuzione rispetto a una scelta che riportava tre categorie possibili (*three-alternative forced-choice* 3AFC).

Nel secondo esperimento è stata seguita una metodologia simile a quella seguita da Bigand (2005), e illustrata nella prima sezione di questo capitolo. Gli stessi esempi di esecuzione del primo esperimento sono stati utilizzati per un'analisi di raggruppamento percettivo da parte dei partecipanti al test; ai soggetti è stato permesso di raggruppare intenzioni espressive simili in accordo con le loro preferenze personali. I dati ricavati da questo esperimento sono stati trattati con analisi multidimensionale, per proiettare le relazioni tra i brani su uno spazio a dimensionalità minore, e con analisi di raggruppamento, per evidenziare particolari schemi relazionali nell'organizzazione percettiva; per finire una *bootstrap analysis* ha confermato la validità dei risultati ottenuti.

4.2.2.2 Risultati e discussione

Vengono ora esposti i risultati degli esperimenti effettuati da Mion, e le interpretazioni e considerazioni fatte in merito.

Le coordinate delle nove intenzioni espressive valutate sono state correlate alle features acustiche ricavate all'inizio dello studio, Figura 4.3, lungo le due

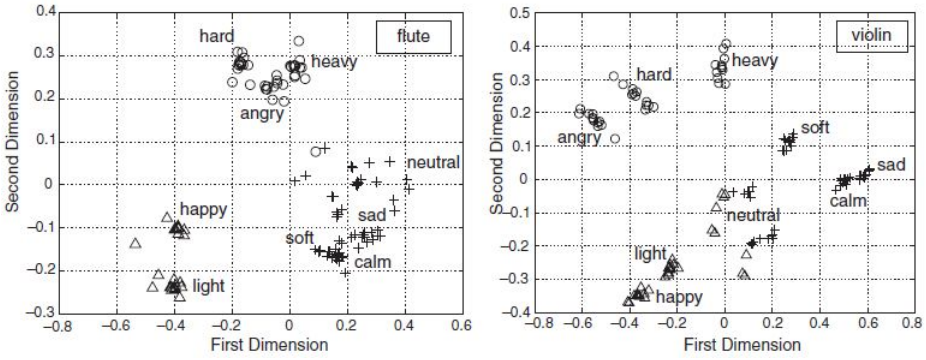


Figura 4.3: Proiezione dell'analisi delle componenti principali per esecuzioni di flauto e violino. (Mion et al., 2010)

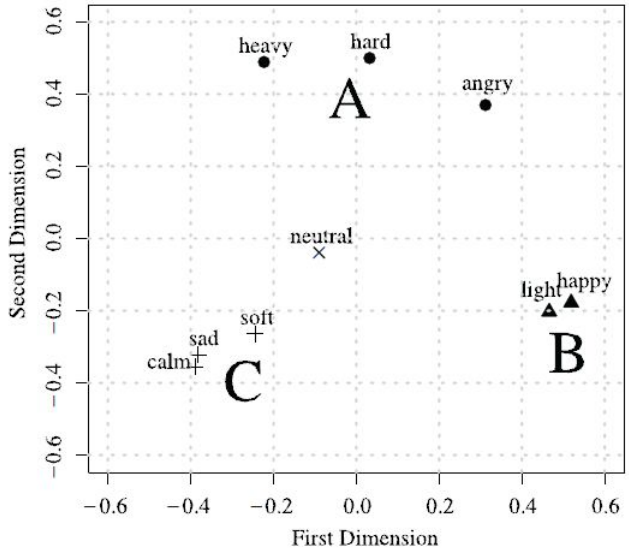


Figura 4.4: Analisi di raggruppamento esperimento 1. (Mion et al., 2010)

dimensioni dello spazio percettivo definito. In questa modo è stato evidenziato come la dimensione 1 sia strettamente correlata con il tempo di attacco, con le note per secondo e la *roughness* (rispettivamente: r=.94, .85, .8), mentre la seconda dimensione con il livello di picco del suono PSL (r=.91). Osservando la configurazione nello spazio percettivo trovato, si nota inoltre come la prima dimensione contrapponga intenzioni espressive da *heavy* a *light* e da *happy* a *sad*; questa opposizione appartiene, rispettivamente, alla dimensione della cinetica nello spazio sensoriale e alla dimensione della valenza nello spazio affettivo; le features che correlano con la prima dimensione, appartengono principalmente alla sfera delle proprietà qualitative dell'articolazione della performance. Valutando invece la seconda dimensione, questa separa intenzioni relative alle dimensioni di *energy* e *activity*; quindi le features che la contraddistinguono sono più di carattere quantitativo, rispetto all'energia, di quanto non lo siano le features relative alla prima dimensione.

La *roughness* è correlata in maniera consistente con entrambe le dimensioni, vista la sua natura di proprietà strutturale del brano; possiede sia aspetti di natura energetica sia qualitativa, infatti è in relazione con la sensazione di sforzo (Serra, 1997). Inoltre, viene notato dagli autori dello studio, come si possa identificare un'opposizione di intenzioni espressive identificate dagli aggettivi *angry/calm* lungo la diagonale dello spazio percettivo; questo si deve al fatto che sono termini dal significato espressivo opposto e che entrambi presentano un andamento rispetto a proprietà qualitative ed energetiche inversamente proporzionale tra loro.

Queste corrispondenze trovano accordo con i risultati presenti in letteratura, soprattutto nei confronti di esperimenti di tipo percettivo messi in atto per studiare e discriminare le differenti emozioni (Sloboda e Juslin, 2001). Sono state confermate, inoltre, le analisi riportate in (Mion e De Poli, 2008) nelle quali le categorie espressive erano poste in posizioni simili rispetto agli stessi parametri acustici.

Valutando inoltre come si relazionano tra loro i cluster definiti in Figura 4.4, si può notare come vi sia una contrapposizione tra il gruppo A e i gruppi B e C. Questa opposizione lungo la seconda dimensione esprime un diverso comportamento a livello energetico tra le intenzioni espressive identificate con *hard/heavy/angry* e le altre prese in considerazione nell'esperimento. Questa contrapposizione si nota sia per il primo esperimento sia per il secondo; inoltre è intuibile una separazione lungo la prima dimensione per quanto riguarda i cluster B e C, che, valutate le features, può essere interpretata come una diversa qualità del suono e una differente articolazione esecutiva.

Mion, quindi, suggerisce due criteri principali di interpretazione dello spazio ricavato: un criterio basato sull'opposizione quantitativa dell'energia, e un secondo relativo alla qualità del suono.

Una ulteriore considerazione è stata fatta in merito ai risultati del secondo esperimento percettivo, dove si richiedeva un raggruppamento personale delle interpretazioni espressive percepite. Ogni gruppo di soggetti ha riconosciuto tre cluster dominanti, sebbene i partecipanti abbiano usato un peso differente nella valutazione stessa; questa diversa interpretazione ha influito identificando una

diversa configurazione dei tre cluster nello spazio percettivo bi-dimensionale, e quindi un differente uso delle due dimensioni principali. Mion suggerisce che i soggetti abbiano usato due differenti metodi: un primo approccio più semplice basato sulla qualità dell'estratto, e un secondo più esperto che ha considerato, oltre alla qualità, anche l'energia espressa nel contenuto musicale.

La rappresentazione teorica dello spazio percettivo, solitamente composta da due dimensioni ortogonali tra loro, in questo studio cambia lasciando il posto a una rappresentazione triangolare, dove per esempio le intenzioni espressive *calm* e *sad* vengono raggruppate insieme. Questa organizzazione a due dimensioni non rende evidente, in parte, la distinzione tra espressioni raggruppate insieme, e quindi manca di una completa distinzione a livello di valenza ed energia. L'inserimento di una terza dimensione, però, limita questo problema di distinzione, riuscendo per esempio a differenziare la categoria *sad* dalla categoria *calm*; aggiungendo questa terza dimensione l'*explained variance* aumenta del 7%, e il modello risulta maggiormente completo. La terza dimensione contribuisce a definire due gruppi principali: un gruppo che vede la separazione tra l'espressione neutra e le altre, e un gruppo che distingue le intenzioni espressive del dominio sensoriale da quelle del dominio affettivo.

Sulla base dei risultati proposti in questa parte, saranno presentati nei prossimi due capitoli esperimenti di tipo percettivo che mirano ad approfondire la conoscenza della struttura dimensionale per il riconoscimento delle intenzioni espressive nella musica.

Capitolo 5

Esperimento 1: brani in modalità maggiore

5.1 Introduzione

In questo capitolo viene presentato e analizzato il primo esperimento percettivo proposto. Lo scopo di questo studio è stato analizzare le intenzioni espressive trasmesse da una selezione di estratti musicali con la caratteristica di essere tutti in tonalità maggiore. Con questo esperimento si cerca di scoprire quali siano le componenti più importanti che discriminano le intenzioni espressive percepite dagli ascoltatori dopo aver eliminato la componente dovuta alla differente tonalità, quest'ultima già identificata come componente fondamentale rispetto al diverso valore di valenza percepita (capitolo 4).

I dati sono stati raccolti nel lavoro di Paganin et al. (2010). Dopo una prima parte del capitolo in cui vengono spiegate le modalità di metodologia nell'esecuzione dell'esperimento, vengono presentate in dettaglio l'analisi dei dati e la discussione dei risultati ottenuti.

5.2 Metodo

La metodologia seguita per questo esperimento segue l'analisi proposta da Bigand (Bigand et al., 2005) e spiegata nel capitolo 4. L'approccio usato è di tipo olistico (sezione 1.2), che comporta test con un grande numero di stimoli e si prefigge lo scopo di valutarne la complessa relazione che esiste tra essi; nel caso specifico, la modalità di esecuzione viene eliminata come fattore discriminante di scelta, rimangono invece tutte le altre features caratteristiche dei brani. L'approccio olistico permette una relazione diretta tra il materiale utilizzato e uno spazio a bassa dimensionalità che descriva la relazione tra i brani; quest'ultimo è calcolato nel corso dell'analisi dei dati e propone risultati importanti per

la comprensione dell'organizzazione mentale degli ascoltatori nei confronti dei diversi brani.

5.2.1 Partecipanti

I partecipanti all'esperimento sono stati in totale 40, suddivisi in 15 di sesso femminile e 25 di sesso maschile. È stata posta particolare attenzione nel differenziare la diversa competenza in ambito musicale: è stata valutata la distinzione tra non musicisti, ossia coloro che non hanno particolari studi in campo musicale ma che ascoltano solamente musica per diletto, e i musicisti, categoria in cui rientrano coloro che hanno almeno cinque anni di formazione a livello musicale, non facendo distinzione tra studi di tipo strumentale o di tecniche di elaborazione del suono. Nel corso dell'analisi vengono indicati con la lettera N i non musicisti e con M i musicisti. L'età dei partecipanti varia dai 20 ai 60 anni, con una media di circa 25 anni.

5.2.2 Materiale

I test sono avvenuti in un laboratorio adibito per lo scopo. Questo ha avuto il vantaggio di far sentire gli utenti più a loro agio in modo da non influenzare i risultati con fattori esterni. La strumentazione utilizzata è di tipo professionale, sia per quanto riguarda i diffusori usati (Genelec 8030A) sia per le cuffie (AKG K501); ai partecipanti è stata lasciata la scelta della modalità di diffusione del suono in base alle loro preferenze, ulteriore attenzione per poter mettere nelle condizioni ottimali il soggetto.

È stato selezionato un campione di 23 estratti musicali presi dal repertorio classico occidentale; i brani hanno in comune il fatto di non avere parti vocali. Alcuni estratti sono comuni alla selezione fatta da Bigand (2005), da dove sono stati esclusi i brani in modalità minore; i brani considerati sono i numeri 1, 4, 5, 6, 11, 13, 14, 15, 21, 22, 23. Tutti gli estratti considerati sono stati scelti per proporre all'ascoltatore un'ampia varietà di intenzioni espressive; inoltre i criteri di scelta hanno compreso anche il differente periodo di collocazione nel repertorio classico occidentale (barocco, classico, romantico) e la diversa strumentazione (solo, musica da camera, orchestrale). Non sono stati scelti estratti di opere famose per evitare l'influenza da parte dell'ascoltatore dovuta a fattori emozionali precedenti all'esperimento. I 23 brani così scelti, della durata media di circa 35 s, sono riportati in Tabella 5.1.

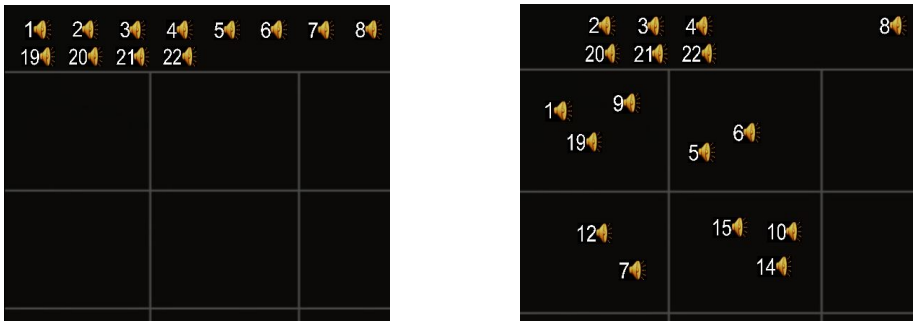
5.2.3 Interfaccia grafica

I file musicali sono stati presentati agli ascoltatori tramite un'interfaccia implementata con il software Pure Data[1], un ambiente di elaborazione del suono in tempo reale (Figura 5.1a). Lo scopo di questa interfaccia è stato di permettere una facile interazione dei partecipanti con l'ascolto dei brani e il semplice

[1] http://puredata.info

#	**Titolo Opera**
1	Strauss - Also sprach Zarathustra
2	Vivaldi - Trio Sonata Do Mayor, RV82 allegro
3	Berlioz - La dannazione di Faust - Ballet des Sylphes
4	Brahms - Violin Concerto, Adagio
5	Scarlatti - Sonata A for Harpsichord, K208
6	Schumann - Tra eumerei, op. 15, no. 7
7	Bizet - Symphony No.1 in C
8	Boccherini - Minuetto
9	Byrd - Galliard
10	Debussy - Claire de lune
11	Liszt - Poeme symphonique
12	Corelli - violin sonata
13	Faure - nocturne op 84 no 8
14	Beethoven - Piano, Sonata 32, mvt 2
15	Mendelssohn - Italian Symphony, mvt 1
16	Handel - Concerto Grosso Op 6
17	Marcello Benedetto - Sonata No.1 in F Major
18	Monteverdi - Prologo - Toccata
19	Haydn - Symphony Bdur - Hob I 105, Andante
20	Bach - Duetto for two utes in G, Allegro
21	Schubert - Valse no. 3, op. 50, D779
22	Beethoven - Symphony. 7, Vivace
23	Monteverdi - Sonata sopra Sancta Maria

Tabella 5.1: Brani in modalità maggiore selezionati per l'esperimento.



(a) Disposizione iniziale delle icone nell'interfaccia. I brani sono associati in modo casuale ai numeri. È possibile ascoltare un brano mettendo il cursore sopra la relativa icona.

(b) Esempio di configurazione durante un esperimento: le icone sono state parzialmente raggruppate a seguito degli ascolti.

Figura 5.1: Interfaccia in Pure Data.

raggruppamento degli stessi; anche questa semplicità di comprensione ed esecuzione del test è stata pensata per non disturbare il soggetto con operazioni poco intuitive che potessero distrarlo dagli scopi dell'esperimento. L'interfaccia è stata creata per permettere l'ascolto dei brani in momenti diversi, sia in fase iniziale e sia in fase di raggruppamento, per evitare che il soggetto non riesca a organizzarsi per cause dovute alla memorizzazione delle intenzioni espressive percepite. L'interfaccia propone ai soggetti una serie di brani evidenziati numericamente da 1 a 23, con l'accortezza che per ogni esperimento l'ordine degli stessi sia casuale; questa casualità evita problemi dovuti alla stessa disposizione dei brani e quindi a una possibile percezione condizionata alla scaletta di ascolto.

Come si può vedere in Figura 5.1a i brani sono rappresentati da una serie di icone a forma di altoparlante. Con il passaggio del cursore del mouse sopra a un brano l'icona viene selezionata (diventa più grande) e inizia la riproduzione del file musicale associato a essa. Il raggruppamento dei brani è possibile trascinando le icone tramite cursore nei quadranti disegnati dalla tabella sottostante (Figura 5.1b). La riproduzione musicale è possibile anche una volta spostate le icone, e non vi è limite nello spostamento. Per finire le configurazioni sono state salvate in un file di testo, con i rispettivi elenchi di ordine dei brani e le durate degli esperimenti.

5.2.4 Procedimento

L'esecuzione dei vari test percettivi è stata fatta in un laboratorio adibito esclusivamente per questo scopo. All'inizio di ogni test ai partecipanti è stata spiegata la modalità di utilizzo dell'interfaccia ed è stato consegnato un foglio con le istruzioni per la corretta esecuzione. È stato sottolineato in particolare che i

brani possono essere ascoltati un numero di volte non definito, possono essere raggruppati provvisoriamente e poi ascoltati nuovamente ed eventualmente spostati in un nuovo gruppo. Non è stato indicato un tempo limite per la durata complessiva del test, ciascun soggetto si è sentito libero di affrontare l'ascolto a seconda della propria preferenza. È stata data la possibilità di scegliere la modalità di diffusione del suono, se tramite cuffie o altoparlanti, in quanto lo spazio riservato ai test era adibito solo per le prove e non vi erano distrazioni esterne.

Come visto nella sezione 1.3 una metodologia congeniale per valutare la similitudine tra brani musicali è l'approccio *object-grouping*, nel quale si richiede di raggruppare le intenzioni espressive percepite per similitudine; è stato usato questo metodo per l'esperimento, ed è stato spiegato ai partecipanti, in particolare, di porre attenzione alle intenzioni espressive percepite, all'esperienza emotiva provata durante l'ascolto: questo per evitare valutazioni che comportino un'organizzazione dei brani su base stilistica o su aspetti puramente strutturali; condizione, questa, più delicata se il partecipante ha una buona cultura musicale, come possono essere i soggetti catalogati come musicisti.

L'esperimento ha avuto una durata in media di 18 minuti, durante i quali i soggetti hanno potuto ascoltare i brani più di una volta, sia in fase di organizzazione e sia in fase di convalida dei gruppi predisposti. È da rilevare come i partecipanti musicisti abbiano avuto meno problemi nella comprensione delle istruzioni dell'esperimento, mentre i soggetti non musicisti hanno avuto la necessità di ulteriori spiegazioni, relative soprattutto alla modalità di percezione sulla quale basarsi per l'organizzazione dei brani.

5.3 Elaborazione dati

L'analisi dei dati è stata suddivisa in quattro fasi: creazione di una matrice di dissimilarità in grado di relazionare i vari brani tra loro in base alle scelte di raggruppamento dei partecipanti; analisi multidimensionale per definire la dimensionalità ottimale per riprodurre le distanze tra i vari oggetti, relativamente alla similitudine che esiste tra loro; analisi di conferma della bontà dei dati raccolti per definire la stabilità del risultato trovato; infine, un'analisi di raggruppamento che permetta di definire il miglior raggruppamento, per similitudine di intenzione espressiva percepita, degli estratti musicali. Questa metodologia è stata precedentemente utilizzata in letteratura per l'elaborazione dei dati relativi ad altri esperimenti (Bigand et al., 2005).

5.3.1 Correlazione matrici dissimilarità

Nella prima fase di analisi è stata calcolata la matrice di dissimilarità totale relativa ai raggruppamenti dei brani fatti durante l'esperimento. La matrice di dissimilarità, chiamata anche matrice delle distanze, descrive la distinzione che esiste tra coppie di oggetti. È una matrice quadrata simmetrica con l'elemento $[i, j]$ rappresentato da un valore che descrive il grado di distinzione tra l'oggetto

Algorithm 5.1 Calcolo della matrice di dissimilarità.

```
##-------------------------------------------------------
##Creazione matrice dissimilarità 23brani * 23brani
##(inizializzo a 0)
dissimilarity<-matrix(0,23,23)
for (k in 1:20){
    for (i in 1:23){
        for (j in 1:23){
            if (scelte[k, i] != scelte[k, j]) {
                dissimilarity[i, j] <- dissimilarity[i, j]+1;
            }
       }
    }
}
##creo le label per la matrice con i valori
label <- c(paste("brano",1:23))
dimnames(dissimilarity) <- list(label,label)
```

i-esimo e l'oggetto *j-esimo*. Gli elementi lungo la diagonale non sono considerati, solitamente vengono posti a zero come in questo caso. La matrice delle distanze riporta i valori relativi alla distinzione tra entità, la matrice di similarità invece descrive la relazione di similitudine tra coppie di oggetti.

Nel caso in oggetto la matrice risultante, definita con A, è una matrice quadrata di dimensione 23, e a ogni riga è associato un brano, come anche per le colonne. Il generico elemento $A[i, j]$ è pari al numero di volte in cui il brano *i-esimo* è stato raggruppato in un gruppo differente del brano *j-esimo*; secondo questa notazione, se $g(i)$ è il gruppo del brano *i-esimo*, allora è possibile definire l'elemento generico della matrice come:

$$A[i,j] = \begin{cases} A[i,j]+1 & g(i) \neq g(j) \\ A[i,j]+0 & altrimenti \end{cases} \tag{5.1}$$

L'elemento generico della matrice, escludendo la diagonale, varia da 0 (i due brani sono sempre risultati nel medesimo gruppo di selezione) a 40 (non sono mai stati valutati come simili per intenzioni espressive); 40 sono i partecipanti all'esperimento. La matrice di dissimilarità è stata prima calcolata separatamente per i dati relativi ai non musicisti e quindi per i musicisti. Viene riportato in Algoritmo 5.1 un estratto del codice in R[2] per il calcolo della matrice.

È stato calcolato successivamente un indice di correlazione tra le due matrici. Per calcolare suddetto indice è stato utilizzato il test di Mantel, che permette di calcolare la correlazione tra matrici di distanze. Questo test considera il caso in cui vengano analizzate due matrici delle distanze tra n oggetti, per un totale di *n(n-1)/2* valori. Per il fatto che le distanze non sono indipendenti una

[2] R è un linguaggio di programmazione e un ambiente per l'elaborazione statistica. È un progetto GNU simile al linguaggio proprietario S. R fornisce una serie di strumenti molto ampia per l'analisi statistica e per l'implementazione grafica dei risultati. Per maggiori informazioni: http://www.r-project.org/index.html

dall'altra – se viene cambiata una posizione di un brano cambiano *n-1* valori – non è possibile calcolare la semplice correlazione che esiste tra le distanze e valutarne il valore statistico. Il test di Mantel evita questo problema sfruttando la permutazione delle righe e delle colonne: dopo aver calcolato il coefficiente di correlazione iniziale, le righe e le colonne vengono permutate casualmente per un numero elevato di volte, e dopo ogni permutazione viene calcolato nuovamente l'indice di correlazione; il significato della correlazione osservata è la proporzione degli indici calcolati per tali permutazioni, infatti se l'ipotesi nulla di non correlazione tra le due matrici fosse vera si avrebbe un incremento o un decremento dell'indice di correlazione. Inoltre il test di Mantel permette il calcolo della correlazione tra matrici di distanze senza nessuna assunzione iniziale rispetto alla distribuzione degli elementi nelle matrici.

Sono stati eseguiti tre test di Mantel per il calcolo della correlazione tra la matrice di dissimilarità associata ai musicisti e la matrice dei non musicisti, e in ogni test è stato utilizzato un differente metodo di calcolo della correlazione tra le coppie di variabili:

- *Pearson product-moment correlation coefficient*
- *Spearman's rank correlation coefficient*
- *Kendall rank correlation coefficient*

Questi indici di correlazione variano da un valore massimo di 1, interpretato come una totale correlazione tra le variabili x e y analizzate, fino a un valore minimo di -1 che significa correlazione negativa; un valore di indice pari a 0 è da interpretarsi come indipendenza tra le variabili. I risultati vedono i seguenti valori calcolati: r=.78, r=.78, r=.66, rispettivamente. La correlazione è quindi buona tra le due matrici, con valore positivo, e quindi per questo motivo è stato possibile calcolare una matrice totale di dissimilarità riguardante tutti i dati dei 40 partecipanti, senza distinzione di categoria.

Questo indice di correlazione elevato tra le matrici di dissimilarità indica inoltre una valutazione piuttosto simile e coerente dei brani da parte delle due categorie di partecipanti. Una correlazione prossima allo zero avrebbe implicato una diversa percezione dell'intenzione espressiva musicale, determinata da una diversa preparazione musicale dei musicisti rispetto ai non musicisti.

5.3.2 *Multidimensional Scaling*

Avendo ottenuto nella prima fase una matrice di dissimilarità totale, è possibile ora procedere con il calcolo dello *scaling* multidimensionale. Si chiama *scaling* multidimensionale (MDS: *MultiDimensional Scaling*) l'insieme di procedure che, partendo da una matrice delle distanze tra n oggetti, trova una configurazione "metrica", ossia rappresentabile geometricamente, delle entità in un numero limitato di dimensioni. Nel caso presentato la tecnica MDS usata è di tipo "non-metrico", ossia assume che le misure di prossimità tra le entità analizzate e le distanze tra i punti sulla configurazione geometrica finale siano in relazione monotòna.

***Multidimensional Scaling* non metrico di Kruskal**

L'analisi delle prossimità mira a rappresentare geometricamente n entità per mezzo di altrettanti punti in modo che le distanze tra punti rispettino, entro un grado di approssimazione tollerabile, i vincoli di partenza. Il modello di analisi assume che nelle prossimità siano latenti q dimensioni e che queste si possano identificare combinando le informazioni rilevate. Ammettendo la possibilità di errore statistico, il modello stabilisce la seguente relazione monotòna tra distanze geometriche finali e prossimità originarie:

$$d_{ij} = R({}_2d_{ij}) + \varepsilon_{ij} \qquad (i \geq j = 1, ..., n) \tag{5.2}$$

dove d_{ij} è una misura ordinale di dissomiglianza tra le entità i e j; ${}_2d_{ij}$ è la distanza euclidea tra i punti i e j nella configurazione finale e *R(.)* indica il rango occupato da ${}_2d_{ij}$ nell'ordinamento crescente delle distanze; ε_{ij} è lo scarto tra il rango della dissomiglianza e la posizione ordinata della distanza. Il procedimento di ricerca della soluzione ottimale è iterativo e la soluzione trovata può essere sottoposta a traslazione e/o rotazione degli assi ortogonali per aiutare l'interpretazione della configurazione.

Kruskal (1964) propone un indice "di dissociazione", indice di *stress*, tra le prossimità iniziali e quelle della soluzione analitica che gode anche della proprietà di invarianza rispetto a variazioni di scala degli assi ortogonali; questo indice è utilizzato inoltre anche dall'algoritmo per la valutazione della configurazione trovata, quando l'indice è accettabile si dice che la configurazione converge e l'analisi multidimensionale termina. Se ${}_2\hat{d}_{ij}$ denota la distanza tra i punti i e j a un certo passo del processo iterativo, e $f(d_{ij})$ è la migliore funzione monotòna delle prossimità iniziali, la dissociazione tra due misure, calcolata sull'intera matrice, è data da:

$$S = \frac{S*}{T*} = \frac{\left\{ \sum_{i,j}^{n} \left[f(d_{ij}) - {}_2\,\hat{d}_{ij} \right]^2 \right\}^{1/2}}{\left\{ \sum_{i,j}^{n} {}_2\hat{d}_{ij} \right\}^{1/2}} \tag{5.3}$$

S è interpretabile come varianza residua, esprime infatti quanta parte della variabilità delle prossimità osservate non è interpretata dalla configurazione ottenuta. Valori per S superiori a .20 sono da valutare come scarso adattamento dei dati alla dimensionalità proposta, e quindi solo configurazioni che diano indice di *stress* con valore minore sono accettabili.

L'analisi MDS è stata calcolata sulla matrice di dissimilarità totale e varie prove sono state fatte per selezionare la dimensionalità ottimale. L'indice di *stress* di Kruskal, in percentuale, è stato 29% con un'unica dimensione, e quindi la soluzione è stata scartata; con due dimensioni il fattore di *stress* è risultato essere 14% e quindi la soluzione è accettabile, e la configurazione supera i

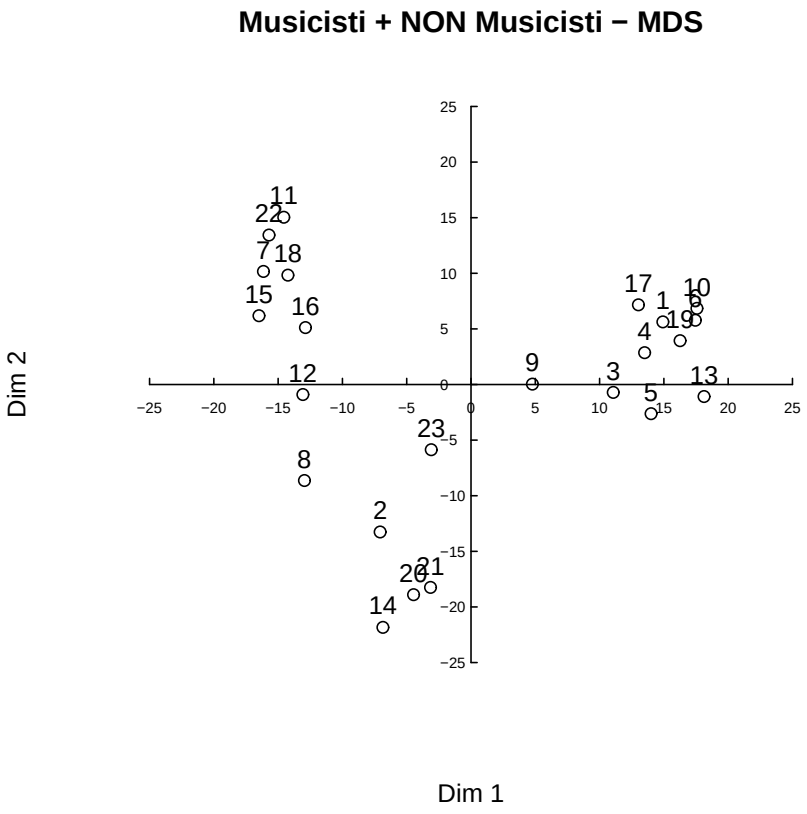


Figura 5.2: Analisi di *scaling* multidimensionale (sono considerati tutti i soggetti, musicisti e non musicisti).

criteri di bontà minimi; prove con ulteriori dimensioni sono state valutate, e le percentuali sono state 8%, 6%, e 4% rispettivamente per i casi con tre, quattro e cinque dimensioni. Notando un decremento non eccessivo dell'indice da una configurazione bidimensionale alle configurazioni con dimensioni aggiuntive è stata quindi selezionata la configurazione a due dimensioni come ottimale (Figura 5.2); i brani valutati vengono dunque disposti in uno spazio con due assi ortogonali, e i punti che li rappresentano sono distanziati in relazione alla similitudine data dai partecipanti all'esperimento (similitudine interpretata dai gruppi formati). Le etichette numeriche assegnate ai punti di Figura 5.2 sono relative ai brani ordinati come in Tabella 5.1.

5.3.3 *Bootstrap Analysis*

Una configurazione è stabile se non risente significativamente della presenza di errori accidentali di rilevazione e della casualità del campionamento delle unità statistiche. Per valutare la stabilità delle configurazioni che si ottengono con l'analisi di *scaling* multidimensionale si possono seguire più vie, la soluzione proposta in questa analisi vede la verifica dell'effetto che ha l'esclusione di una parte delle informazioni iniziali sul risultato; se la configurazione rimane la stessa, o comunque dentro a determinati intervalli di confidenza, significa che le dimensioni trovate sono applicabili all'intera popolazione in esame.

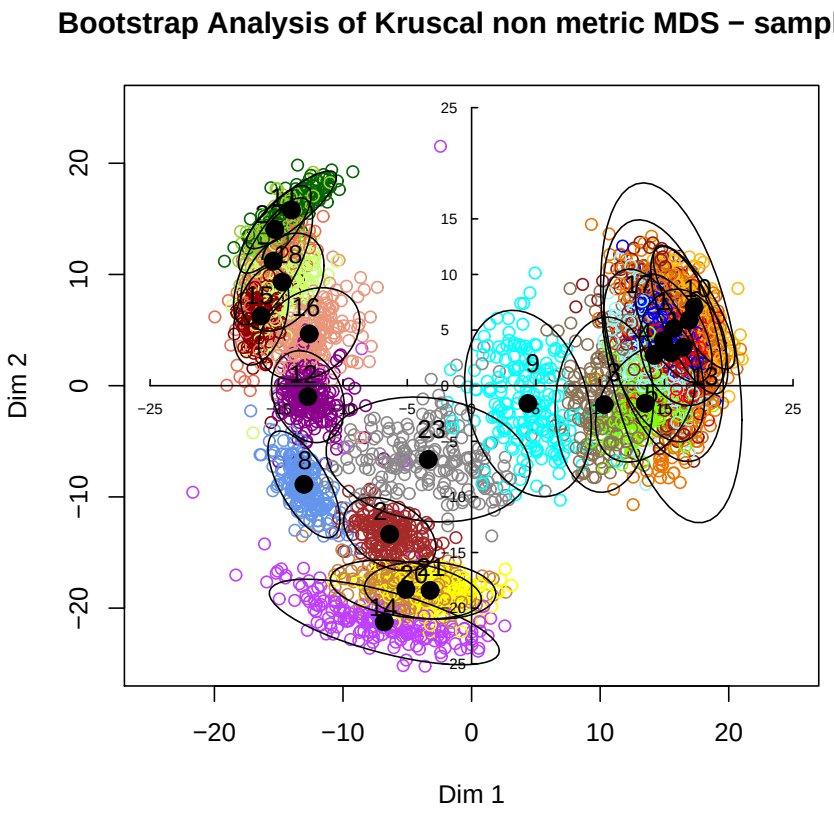
La metodologia del *bootstrapping* serve per testare l'affidabilità di un dataset, in questo caso di una configurazione di punti; l'obiettivo di questo tipo di analisi è di verificare la coerenza dei dati raccolti. Per questa analisi, quindi, lo scopo è di verificare la presenza di eccessive differenze tra i raggruppamenti dei diversi partecipanti, le quali potrebbero influire negativamente sulla configurazione trovata dall'analisi di *scaling* multidimensionale.

I passi principali di un'analisi *bootstrap* sono:

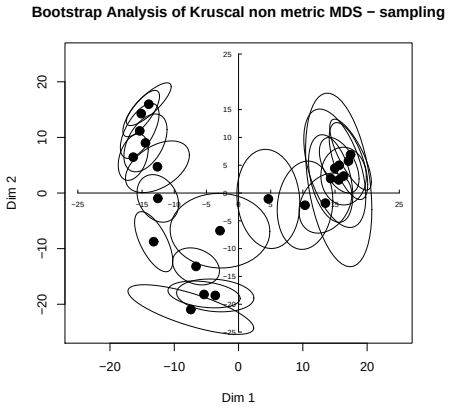
- campionamento casuale dei dati iniziali con reinserimento (RSWR: *Random Sample With Replacement*) e calcolo della statistica;
- iterazione del campionamento casuale con relativa statistica per ogni nuova popolazione di dati;
- confronto dei risultati derivanti dalle diverse iterazioni;
- calcolo, quantitativo o qualitativo, di un intervallo di confidenza adeguato per la *bootstrap analysis.*

Nel caso specifico viene utilizzata la strategia definita *case resampling* con l'implementazione dell'algoritmo Monte Carlo, che determina il campionamento dei dati in modo casuale; questa procedura segue i punti elencati sopra, con il vincolo che il numero di valori del campionamento sia uguale alla dimensione dell'insieme di dati di partenza.

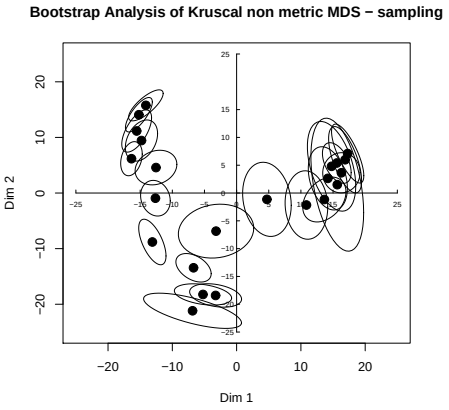
Relativamente all'analisi dei dati di questo esperimento è stato eseguito un campionamento casuale con reinserimento per ottenere un nuovo insieme di dati pari a 40 valutazioni (quanti i partecipanti al test); a questo nuovo insieme di



(a) *Bootstrap analysis* relativa all'analisi MDS di Figura 5.2 per tutti i partecipanti con ellissi di confidenza al 95% per tutti i brani.



(b) Ellissi di confidenza al 95%.



(c) Ellissi di confidenza all' 85%.

Figura 5.3: *Bootstrap analysis* con ellissi di confidenza. Sono proposte le soluzioni per tutti i partecipanti con la visualizzazione dei valori delle 200 iterazioni (a), e con i soli ellissi di confidenza al 95% (b) e all'85% (c).

dati è stata riproposta un'analisi di *scaling* multidimensionale con due dimensioni imposte, sfruttando il risultato descritto nella sezione precedente. Questa elaborazione è stata compiuta iterativamente per 200 volte. Come risultato finale è stata valutata qualitativamente la stabilità della configurazione trovata in precedenza attraverso la sovrapposizione delle 200 configurazioni ricavate da altrettante iterazioni. Inoltre è stato calcolato per ogni brano un ellisse di confidenza che contenga la percentuale di valori desiderata; in Figura 5.3a viene presentato il risultato della sovrapposizione delle analisi con relativi ellissi di confidenza al 95%; nella parte inferiore vengono proposti i soli ellissi senza i punti derivati dalle analisi MDS, in particolare in Figura 5.3b la percentuale di confidenza è al 95% mentre in Figura 5.3c la percentuale è dell'85%.

Per ottenere questa rappresentazione è stata necessaria una rotazione degli assi in alcune iterazioni, infatti come spiegato nella sezione precedente, l'analisi di *scaling* multidimensionale può ruotare gli assi per una migliore disposizione dei risultati, fatto che impedirebbe la corretta visualizzazione di confronto per la *bootstrap analysis*.

5.3.4 *Cluster Analysis*

A seguito della fase di costruzione di uno spazio ottimale a bassa dimensionalità che mantenga le distanze tra i punti dei brani, è stata eseguita un'analisi di raggruppamento con lo scopo di creare gruppi con entità omogenee (*cluster analysis*). L'analisi di raggruppamento si distingue dall'analisi fattoriale perché la prima è pertinente al *clustering* di entità, mentre la seconda è adatta per lo studio delle relazioni tra le variabili. Inoltre, l'analisi di raggruppamento ha il vantaggio che non richiede la forma delle relazioni tra le variabili, e quindi si adatta anche a un caso, come quello in esame, in cui non vi è conoscenza iniziale delle relazioni tra brani. In questa analisi sono proposte e messe a confronto due tipi di *cluster analysis*: l'analisi *k-means* e la *PAM clustering*.

Metodo di analisi *k-means*

Questo metodo rientra nella categoria delle analisi di raggruppamento di tipo gerarchico divisivo, che si fondano sulla divisione dell'insieme di entità sulla base di un attributo dicotomico per volta, o di tutti gli attributi a un tempo. I criteri divisivi sono più generali di quelli agglomerativi perché permettono la formazione di un numero qualsiasi di sottogruppi da un gruppo genitore. Partendo da una situazione in cui le n unità fanno parte di un gruppo unico, in n-1 passi l'algoritmo riesce a formare n gruppi, ognuno formato da una unità (caso estremo).

Il metodo proposto da MacQueen (1967) e noto come *k-means* è un metodo che si basa sulla distanza tra centroidi:

- al primo passo si effettua una suddivisione in due gruppi sulla base della combinazione delle unità che minimizza la devianza interna ai gruppi;

# cluster	2	3	4	5	6	7	8
PAM	.27	.29	.24	.26	.25	.24	.22
k-means	58 %	89 %	92 %	94 %	96 %	97 %	97 %

Tabella 5.2: Analisi di raggruppamento: valori calcolati.

- a ogni successivo passo, individuato il gruppo che ha la massima devianza interna, detta anche devianza di un elemento dal centroide, la suddivisione in un numero k di partizioni delle n unità del gruppo si effettua provando tutte le possibili combinazioni, individuando quella che minimizza la funzione:

$$ {}_kD^2 = \sum_{g}^{G} \sum_{i}^{p} ({}_g\bar{x}_i - {}_g\,\bar{x})^2 \tag{5.4} $$

  dove ${}_g\bar{x}_i$ è il valor medio della variabile i nel sottogruppo g, detta anche funzione di distanza non ponderata, e p rappresenta il numero di entità all'interno di una data partizione.

PAM clustering

Comparato con il metodo di analisi *k-means*, il metodo di partizionamento secondo *medoids* (PAM) comporta le seguenti caratteristiche:

- opera sulla matrice di dissimilarità di un insieme di dati;
- è più robusto, perché minimizza la somma delle dissimilarità invece della somma dei quadrati delle distanze euclidee;
- non necessità a priori la determinazione della quantità di *cluster* da determinare, ma ne valuta di volta in volta la soluzione migliore.

Come primo passo l'algoritmo PAM definisce k oggetti rappresentativi, chiamati *medoids*; un *medoid* può essere definito come un oggetto di un *cluster* la cui dissimilarità media rispetto a tutti gli oggetti nel *cluster* è minima, in questo modo esso sarà il punto più centrale di un dato *dataset*. Dopo aver definito questi *medoids* ogni oggetto dell'insieme di partenza viene assegnato al medoide più vicino; si cerca di minimizzare la funzione obiettivo iterando questo primo passo, in modo che la somma delle distanze di ogni entità dal proprio medoide sia la minore possibile.

Il metodo di raggruppamento PAM è stato eseguito sulla matrice di dissimilarità; questa soluzione ha il vantaggio di definire la soluzione di raggruppamento migliore a prescindere dall'analisi di *scaling* multidimensionale che verrà fatta. In Tabella 5.2 sono elencati i risultati dell'analisi fatta con il metodo PAM: i valori indicano il grado medio di appartenenza di un oggetto a un *cluster*,

quindi un valore più alto possibile è da preferirsi perché indica una soluzione migliore di raggruppamento. Si nota facilmente come la soluzione migliore sia rappresentata da 3 *cluster*.

Come detto in precedenza, l'utilizzo della tecnica del *k-means* comporta una definizione a priori del numero di *cluster* da formare, e quindi sono state fatte più analisi di raggruppamento per stabilire quale soluzione meglio si adatti allo spazio bidimensionale ricavato dall'analisi MDS. Sempre in Tabella 5.2 si vedono elencati i valori di *explained variance* rispetto a soluzioni che vanno da soli due gruppi formati a otto. È evidente come anche in questo caso la soluzione ottimale è la composizione di tre gruppi, infatti con un numero superiore di gruppi il valore aumenta ma di poche unità per ogni dimensione successiva, e il valore percentuale con due gruppi non è accettabile.

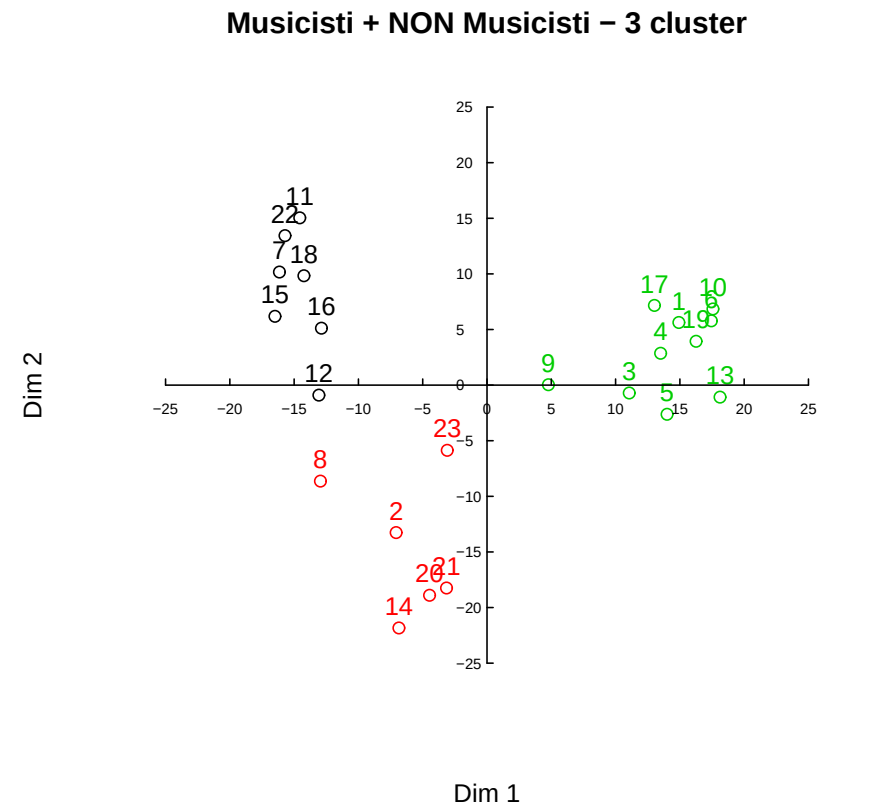
Infine, si è verificato che i gruppi formati dai due metodi siano gli stessi. La Figura 5.4 presenta la soluzione di raggruppamento comune a entrambi gli approcci. Inoltre è stata eseguita la stessa analisi anche per i dati separati per categorie, per scoprire eventuali brani valutati in maniera differente e/o configurazioni di raggruppamento diverse.In Figura 5.4(a) viene proposta la soluzione ottimale con 3 gruppi per i non musicisti e in Figura 5.4(b) la soluzione per i musicisti. Si può osservare una discrepanza di formazione nei gruppi, che verrà analizzata nella sezione 5.5.

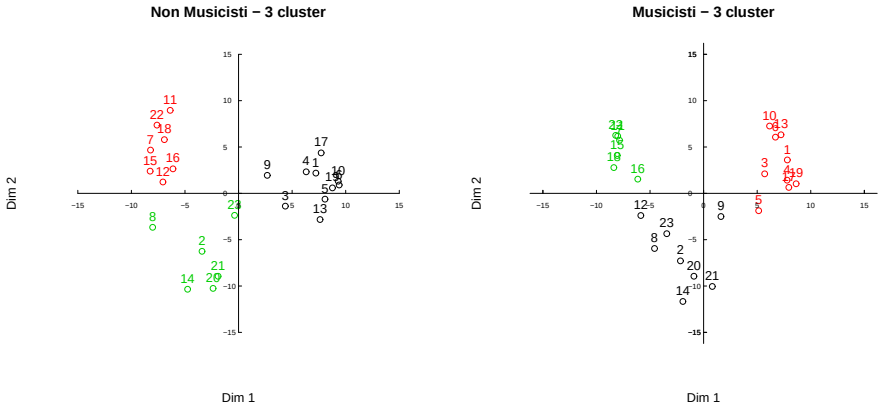
5.4 Estrazione features

Per permettere una corretta analisi delle risposte dei partecipanti è stata necessaria l'estrazione di una serie di features caratteristiche dei brani considerati. Le features calcolate appartengono a un insieme di caratteristiche importanti per la discriminazione delle diverse intenzioni espressive (Sloboda e Juslin, 2001), e sono state usate in vari lavori presenti in letteratura sia per la classificazione dello stile musicale (Dannenberg et al., 1997) sia per la determinazione dei contenuti espressivi nelle performance musicali (Friberg et al., 2002; Mion e De Poli, 2008).

Le features calcolate sono sia di tipo locale e sia riguardanti interi eventi. Le features a livello locale sono calcolate usando frame senza overlap di 46 ms di lunghezza, e poi sono stati considerati i valori medi valutati su finestre della durata di 4 s con overlap di 3.5 s. La particolare taglia della finestra permette di includere una quantità sufficiente di eventi, e corrisponde in parte alla dimensione della memoria ecoica[3]. Inoltre sono state valutate due features a livello di eventi, che sono l'attacco e l'onset; il calcolo per le caratteristiche sugli eventi ha riguardato solo il valore derivante da finestre di 4 s con 3.5 s di sovrapposizione.

[3]Memoria di tipo sensoriale, facente parte della memoria a breve termine (STM: *Short Term Memory*). Questo tipo di memoria è in grado di collezionare una grande quantità di informazioni di tipo auditivo che vengono mantenute solo per un breve periodo di tempo (3-4 s)(Snyder, 2000).





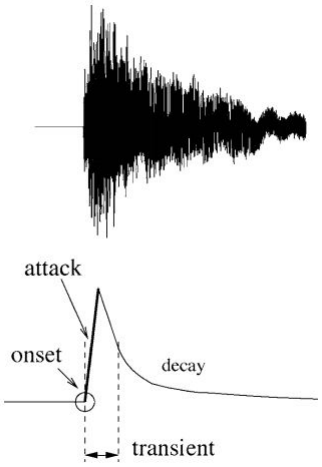
(a) *Cluster analysis* per i non musicisti. (b) *Cluster analysis* per i musicisti.

Figura 5.4: Grafici relativi all'analisi di raggruppamento.

Le features per tutti i brani analizzati sono state estratte utilizzando degli script Matlab[4]. Nello specifico è stata utilizzata la MIRtoolbox[5], che offre una serie di funzioni scritte in Matlab dedicate all'estrazione da file audio di features relative al ritmo, alla struttura del segnale, ecc. Questa toolbox è stata progettata e tuttora supportata dal dipartimento di musica dell'università finlandese di Jyväskylä. Le features calcolate per questo esperimento sono in totale 12.

- **Tempo**. Contraddistingue la velocità della performance; molti degli estratti considerati per l'esperimento hanno una struttura polifonica complessa e quindi non esiste una tecnica automatica per il calcolo di questa features. Per questo il tempo di ogni brano è stato stimato manualmente da un esperto.

- **Attacco**. Modalità di inizio di una frase musicale. Viene valutato come il tempo che intercorre prima di raggiungere il picco dovuto a un evento musicale nell'inviluppo del segnale (da una soglia minima di 2% dell'ampiezza massima del picco al raggiungimento di quest'ultima).



- **Brightness**. Misura della quantità di energia del segnale limitata alla banda di frequenze superiori ai 1000 Hz. Il valore è espresso con un indice da 0 a 1.

- **Centroide**. Un importante descrizione della forma di una distribuzione può essere ottenuta dal calcolo dei suoi momenti; il primo momento è il centro geometrico (centroide) di una distribuzione ed è una misura di tendenza centrale di una variabile casuale. Il centroide spettrale calcolato

[4]MATrix LABoratory è un ambiente per il calcolo numerico e l'analisi statistica. Computazionalmente è più veloce di un tradizionale linguaggio di programmazione come possono essere C, C++ e Fortran. http://www.mathworks.com/products/matlab/

[5]Per maggiori informazioni: https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe-/materials/mirtoolbox

sullo spettro del segnale musicale definisce una misura di luminosità del suono.

- **Onsets**. Un metodo alternativo per determinare una misura simile al tempo è il calcolo della *onset detection curve;* questa curva evidenzia i successivi picchi di energia nell'inviluppo del segnale corrispondenti a successivi battiti. Viene utilizzato un *peak tracking* in grado di stimare la posizione delle note.

- **RMS**. L'energia globale di un segnale x può essere calcolata considerando la radice della media del quadrato dell'ampiezza (*Root Mean Square*):

$$x_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{5.5}$$

- **Rolloff**. Frequenza di soglia sotto la quale è compreso l'85% dell'energia totale del segnale; è relazionato alla brightness del segnale.

- **Roughness**. Stima del grado di dissonanza tra due sinusoidi, a seconda del rapporto tra la loro frequenza. La roughness totale per un suono complesso può essere calcolata determinando i picchi dello spettro, e prendendo la media di tutte le dissonanze tra tutte le coppie di picchi possibili. Il modello di stima utilizzato è proposto da Vassilakis (2001), in cui avendo due sinusoidi f_1 e f_2 e le relative ampiezze A_1 e A_2 è possibile stimare la roughness con la seguente formula:

$$R\left(f_1, f_{2,} A_1, A_2\right) = \left(A_1 * A_2\right)^{0.1} * 0.5 \left(\frac{2A_2}{A_1 + A_2}\right)^{3.11} *$$

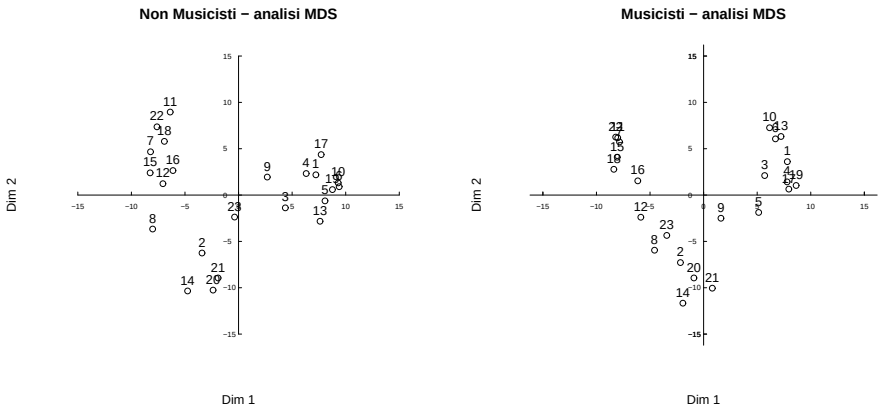
$$* \left[\exp^{-b_1 s(f_2 - f_1)} - \exp^{-b_2 s(f_2 - f_1)}\right]$$

  dove $b_1 = 3.5$, $b_2 = 5.7$, $s = \frac{x*}{s_1 f_1 + s_2}$, $x* = .24$, $s_1 = .02$ e $s_2 = 18.9$.

- **Spectral Ratio**. Viene calcolata la parte di spettro relativamente a due bande di frequenze: SRh indica la la banda più alta, con frequenze che sono maggiori di 1805 Hz; SRm indica la banda compresa tra le frequenze 534 e 1805 Hz. Rappresenta un indice di come viene distribuito lo spettro del segnale rispetto alle varie frequenze.

- **Spectralflux**. Distanza media tra lo spettro calcolato su frame successivi.

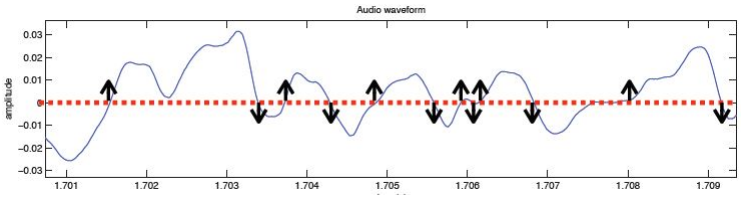
- **Zerocross**. È un indicatore della rumorosità del segnale, e consiste nella somma del numero di volte in cui il segnale cambia di segno (passaggio attraverso l'asse delle ascisse in uno dei due versi).



(a) MDS non musicisti.

(b) MDS musicisti.

Figura 5.5: MDS per le due categorie di partecipanti.



5.5 Discussione risultati

Vengono ora riportati i risultati ottenuti nell'esperimento e discussi in modo da comprendere la bontà dei dati rilevati e le informazioni che comportano.

La correlazione positiva evidenziata dall'analisi delle matrici di dissimilarità dei musicisti e dei non musicisti indica che non vi sono differenze molto spiccate nell'organizzazione mentale delle intenzioni espressive da parte delle due categorie di partecipanti. In particolare, come viene valutato in seguito, l'indice di correlazione potrebbe essere ulteriormente maggiore di quello calcolato (r=.78): risultano esserci infatti alcuni brani di difficile catalogazione da parte dei soggetti.

L'analisi di *scaling* multidimensionale propone una soluzione ottima con struttura bidimensionale. Questo risultato è di particolare importanza perché si ribadisce il modello tridimensionale visto in letteratura (sezione 2.2); escludendo uno dei fattori principali dell'asse della valenza, com'è risultata essere la modalità, viene messa in risalto la struttura rimanente che rispecchia chiaramente la seconda dimensione del modello, l'attivazione, e una terza dimensione non identificabile.

In Figura 5.5 vengono riportate le analisi MDS calcolate per entrambe le categorie separatamente. L'aspetto che risulta essere più evidente è un maggiore accorpamento dei brani nell'analisi dei musicisti; questo significa che i musicisti sono stati più coesi nelle associazioni delle intenzioni espressive, hanno avuto un metodo organizzativo più omogeneo, evidentemente dovuto alla maggiore cultura musicale posseduta che ne influenza la valutazione. Inoltre si notano alcune differenze, seppure non troppo elevate, per la classificazione di alcuni brani, tra cui i numeri 3, 8, 9, 12, 23; questi brani vengono associati dall'analisi di *scaling* a posizioni diverse nei due casi, fatto che comporta un possibile cambiamento di gruppo di appartenenza a seguito dell'analisi di raggruppamento, presentata più avanti. I brani che presentano ua variabilità maggiore contengono una strumentazione non omogenea e hanno caratteristiche dissimili a livello acustico.

L'analisi di *bootstrapping* ha permesso di validare la bontà della configurazione MDS. Per avere una buona stabilità nell'organizzazione ottenuta occorre avere una dispersione non eccessivamente elevata dei valori delle iterazioni della *bootstrap analysis*. La Figura 5.3a mostra un'ottima stabilità nelle scelte, con alcuni casi limite, rappresentati dai numeri 3, 9, 13, 14, 23, che trovano discordanza nella valutazione dei partecipanti. Non vi è una netta classificazione in base alle features per questi brani, ma si può notare in tutti dei valori bassi di centroide, rolloff e zerocross; questi valori indicano brani con poca rumorosità (valore basso di zerocross), e con una distribuzione dello spettro sulle basse frequenze della banda del segnale audio. Soprattutto la poca definizione dell'estratto a livello strutturale, con un armonia elevata, può aver portato valutazioni diverse da parte dei soggetti, con una certa difficoltà nell'identificare l'intenzione espressiva percepita. In generale l'intervallo di confidenza al 95% presentato in Figura 5.3a per l'analisi di *bootstrapping* dei vari brani corrisponde a una buona rappresentazione qualitativa della stabilità nella valutazione degli stessi da parte dei partecipanti, con solo qualche eccezione, come già menzionato.

Considerando l'analisi di tutti i dati raccolti senza distinzione di categoria, l'analisi di raggruppamento ha messo in evidenza 3 gruppi principali in cui possono essere raggruppati i brani (Figura 5.4). La soluzione evidenziata per le due categorie separate, musicisti e non musicisti, ha proposto ugualmente tre *cluster*, con gli stessi brani nei vari gruppi a parte due eccezioni, i brani numero 9 e 12, che sono stati assorbiti dal gruppo inferiore del piano. Questa diversa distribuzione è una diretta conseguenza dell'analisi di *scaling* multidimensionale; unica feature con valori superiori alla media che hanno in comune questi due brani è il tempo di attacco elevato.

È interessante analizzare la variabilità delle features dei brani rispetto alla collocazione degli stessi lungo le due dimensioni del piano. In Tabella 5.3 è possibile vedere una serie di valori derivanti dall'analisi di correlazione tra le singole features con le due dimensioni principali. Si può notare come vi siano valori elevati di correlazione solo in relazione alle features acustiche, e non siano trovate invece relazioni sufficienti con le due features basate sugli eventi. Per la prima dimensione si ha un'alta correlazione negativa con il tempo ($r=-.71$, $p<.001$), con la brightness ($r=-.70$, $p<.001$); questo risultato conferma la seconda dimensione già studiata in letteratura e associata all'*activity* di un brano.

Tabella 5.3: Correlazione dimensioni-features estratte (** $.001 < p < .01$, *$p < .001$).

#	tempo	attacks	brightness	centroid	onsets	rms
dim 1	-.71*	.23	-.70*	-.52	.14	-.31
dim 2	-.19	.14	.40	.47	-.07	-.08
	rolloff	roughness	SRh	SRm	spectralflux	zerocross
dim 1	-.61**	-.30	-.57**	.44	-.68*	-.67*
dim 2	.42	-.11	.29	-.50	-.04	.35

Per quanto riguarda la seconda dimensione vi sono alcune features spettrali con valori piuttosto significativi, ma non emerge un aspetto dominante per la sua caratterizzazione. Questo risultato ha indotto la realizzazione di un ulteriore esperimento di percezione con l'intento di analizzare in dettaglio questa ultima dimensione dello spazio delle intenzioni espressive nella musica (vedi capitolo 6).

Ripetuti ascolti eseguiti, dall'autore di questo lavoro e da esperti, sui brani organizzati secondo l'analisi MDS ha portato alla ulteriore conferma che l'asse delle ascisse è correlato al metronomo inverso, con estratti che variano da allegro a calmo.

5.5.1 Confronto con risultati precedenti

In questa sezione verrà fatta un'analisi di confronto tra le informazioni ricavate in questo esperimento e i risultati presenti in letteratura e proposti da Bigand (sezione 4.1).

Alcuni dei brani utilizzati per questo esperimento sono stati presi dalla lista di estratti considerata da Bigand per i propri esperimenti (Figura 4.1). I brani numero 1, 4, 5, 6, 19 sono inseriti per questo esperimento nel gruppo a destra; i brani 11, 15 e 22 sono presenti nel gruppo in alto a sinistra; i brani 14, 20 e 21 sono nel *cluster* in basso. La disposizione di questi brani nei vari gruppi è associabile ai risultati ottenuti da Bigand, in quanto escludendo i brani in tonalità minore dalla sua lista, sono stati considerati solo gli estratti con un'alta connotazione di valenza; quindi il *cluster* più a destra in questo esperimento raggruppa brani con alta valenza e basso *arousal*, il *cluster* in alto a sinistra rappresenta i brani catalogati da Bigand con alta valenza e alto *arousal*, mentre il gruppo più in basso è identificabile con alto livello di valenza ma poca discriminazione a livello di *arousal* (Figura 5.4).

I risultati sono quindi consistenti con la struttura ricavata dall'analisi fatta da Bigand (2005).

Capitolo 6

Esperimento 2: analisi terza dimensione

6.1 Introduzione

In questo capitolo viene proposto un secondo esperimento di tipo percettivo; questa prova è la conseguenza dei risultati ottenuti nell'esperimento discusso nel capitolo precedente. Riprendendo le informazioni ricavate nel capitolo 5, è stata evidenziata una struttura bidimensionale per l'organizzazione delle emozioni della musica, dopo aver eliminato uno dei fattori di maggiore correlazione del primo asse che è la modalità di esecuzione; inoltre è stata provata una forte correlazione (negativa) tra il metronomo degli stimoli musicali e il primo asse della Figura 5.2. Avendo queste informazioni e volendo analizzare in maggior dettaglio questa struttura a bassa dimensionalità è interessante considerare un esperimento nel quale vengano proposti solamente brani con la stessa modalità e con lo stesso tempo di esecuzione.

Questo esperimento, come il precedente, non è mai stato provato in letteratura, e rappresenta una buona soluzione per cercare di definire al meglio il modello tridimensionale delle emozioni percepite tramite la musica; in particolare si cercano correlazioni con le caratteristiche del segnale audio in grado di descrivere la terza dimensione identificata in letteratura (cfr. sezione 2.2).

6.2 Metodo

La metodologia seguita per l'esecuzione di questo esperimento segue i passi dell'analisi proposta da Bigand (2005) e già utilizzata per il primo esperimento proposto nel capitolo precedente. L'approccio rimane di tipo olistico (cfr. sezione 1.2) in quanto, nonostante vengano eliminate alcune componenti di fondamentale importanza per la discriminazione delle intenzioni espressive dei brani, si cercano di mettere in luce relazioni tra una serie di caratteristiche del segnale

audio e la disposizione che sarà ricavata dall'analisi; il numero di features resta quindi numeroso nonostante questa ipotesi iniziale di limitazione. Le caratteristiche del segnale considerate restano molte per il motivo che le dimensioni meno definite della struttura non correlano fortemente solo con una caratteristica ma saranno descritte da una serie di correlazioni con valori non elevati, e quindi è necessario prendere in considerazione un'ampia scelta di variabili acustiche.

6.2.1 Partecipanti

I partecipanti all'esperimento, come per l'esperimento precedente, sono stati in totale 40 suddivisi in 12 di sesso femminile e 28 di sesso maschile. È stata posta, anche in questa occasione, particolare attenzione nel differenziare la diversa competenza in ambito musicale: è stata valutata la distinzione tra non musicisti, ossia coloro che non hanno particolari studi in campo musicale ma che ascoltano solamente musica per diletto, e i musicisti, categoria in cui rientrano coloro che hanno almeno cinque anni di formazione musicale, non facendo distinzione tra studi di tipo strumentale o di tecniche dell'elaborazione audio. Sono stati considerati quindi 20 non musicisti e 20 musicisti, quasi tutti, quest'ultimi, facenti parte del Conservatorio Statale di Musica "C. Pollini" di Padova. L'età dei partecipanti varia dai 20 ai 45 anni, con una media di circa 25 anni.

6.2.2 Materiale

L'esperimento è avvenuto in parte in un laboratorio presso il complesso Dei/O della facoltà di Ingegneria dell'Università di Padova e in un ufficio presso la sede staccata del conservatorio Pollini; entrambi queste disposizioni sono state adibite esclusivamente per gli scopi dei test. Questo ha avuto il vantaggio di far sentire gli utenti più a loro agio in modo da non influenzare i risultati con fattori esterni. La strumentazione utilizzata è di tipo professionale, sia per quanto riguarda i diffusori (Genelec 8030A) sia per le cuffie (AKG K501); ai partecipanti è stata lasciata la scelta della modalità di diffusione del suono in base alle loro preferenze, ulteriore attenzione per poter mettere nelle condizioni ottimali il soggetto.

È stato selezionato un campione di 22 brani musicali presi dal repertorio classico occidentale; i brani hanno in comune l'assenza di parti vocali. Alcuni estratti sono stati presi dalla selezione fatta da Bigand (2005) e sono i numeri 13, 14, 22; i brani 8 e 12 era stati considerati anche per l'analisi del primo esperimento. Tutti gli altri estratti sono stati scelti cercando di rispecchiare le ipotesi di partenza: modalità maggiore e stesso metronomo. Sono stati individuati quindi un totale di 22 pezzi musicali con tempo di esecuzione tra 96 e 108 bpm[1]; dopo un primo ascolto è stato evidenziato come fosse possibile ancora riconoscere la differenza di tempo tra un brano a 96 bpm e un brano a 108 bpm. Per avere un indicazione di tempo omogenea si è pensato di standardizzare i

[1]Battiti Per Minuto: è un'unità di misura di frequenza, viene usata principalmente in musica per indicare il valore del metronomo.

brani a un metronomo comune di 104 bpm (per i brani che già non erano eseguiti con questo tempo); questo processo ha visto la modifica tramite software del metronomo degli estratti, cambiamento che è stato inferiore all'8% e che quindi mantiene le caratteristiche dei vari brani senza influire negativamente su altre componenti del segnale.

Tutti gli estratti considerati sono stati scelti per proporre all'ascoltatore una ampia varietà di intenzioni espressive; inoltre i criteri di scelta hanno compreso anche il differente periodo di collocazione nel repertorio classico occidentale (barocco, classico, romantico) e la diversa strumentazione (solo, musica da camera, orchestrale). Non sono stati scelti estratti di opere famose per evitare l'influenza da parte dell'ascoltatore dovuta a fattori precedenti all'esperimento. I 22 brani così scelti, con durata media di circa 26 s, sono riportati in Tabella 6.1.

6.2.3 Interfaccia grafica

I file musicali sono stati sottoposti agli ascoltatori tramite un'interfaccia implementata con il software Pure Data, simile all'interfaccia presentata nella sezione 5.2.3 per il primo esperimento. Lo scopo di questa interfaccia è stato di permettere una facile interazione dei partecipanti con l'ascolto dei brani e il semplice raggruppamento degli stessi. L'interfaccia è stata implementata per permettere l'ascolto dei brani in momenti diversi, sia in fase iniziale e sia in fase di raggruppamento, per evitare che il soggetto non riesca a organizzarsi per cause dovute alla memorizzazione delle intenzioni espressive percepite. L'interfaccia propone ai soggetti una serie di brani evidenziati numericamente da 1 a 22, con l'accortezza che per ogni esperimento l'ordine degli stessi sia casuale; questa casualità evita problemi dovuti alla stessa disposizione dei brani e quindi a una possibile percezione condizionata alla scaletta di ascolto.

I partecipanti possono organizzare in una griglia le icone relative ai brani ascoltati, in modo da creare gruppi basati sulle intenzioni espressive percepite da ciascuno (Figura 5.1b). Anche in questa prova è stata data la possibilità di ascolto durante tutte le fasi dell'esperimento e di modifica dei gruppi creati.

6.2.4 Procedimento

Come introdotto in precedenza, gli esperimenti si sono svolti in spazi adibiti appositamente per queste prove, senza interferenze per cause esterne; all'inizio di ogni test ai partecipanti è stata spiegata la modalità di utilizzo dell'interfaccia ed è stato consegnato un foglio con le istruzioni per la corretta esecuzione. È stato sottolineato in particolare che i brani possono essere ascoltati un numero di volte non definito, possono essere raggruppati provvisoriamente e poi ascoltati nuovamente ed eventualmente spostati in un nuovo gruppo. Non è stato indicato un tempo limite per la durata complessiva del test, ciascun soggetto è lasciato libero di affrontare l'ascolto a seconda della propria preferenza. È stata data la possibilità di scegliere la modalità di diffusione del suono, se tramite cuffie o altoparlanti, in quanto lo spazio riservato ai test era adibito solo per le prove e non vi erano distrazioni esterne.

Tabella 6.1: Brani in modalità maggiore e con lo stesso metronomo selezionati per l'esperimento. I criteri di scelta hanno considerato il differente periodo di collocazione nel repertorio classico occidentale (barocco, classico, romantico) e la diversa componente strumentale (solo, musica da camera, orchestrale).

#	**TitoloOpera**
1	Beethoven - Symphony 7
2	Beethoven - Piano, Sonata 32, mvt 2
3	Bach - Duetto for two flutes in G
4	Vivaldi - Trio Sonata Do Mayor, RV82 allegro
5	Beethoven - Andante con variazioni for Mandolin and Piano in D Major, WoO 44 No.2
6	Boccherini - Minuetto
7	J. Brahms - Violin Concerto in D major - I. Allegro non troppo (Perlman_Giulini) Part 1
8	Carolan's Concerto (Carolan Five Tunes by the Irish Harper) - Da Camera
9	Corelli - violin sonata
10	Galuppi - Sonata in do maggiore - Andante (1_3)
11	Haendel - Concerto a Due Cori 3.6 Allegro
12	Haendel - Zadok the Priest HWV 258
13	Haydn - Trumpet Concerto in E flat major (part 2)
14	Joseph Martin Kraus - Piano sonata in E-flat major, VB195 - Andante con variazione
15	Monteverdi - Prologo - Toccata
16	Mozart - Concerto No.27 in B flat, 3rd mov
17	Mozart - Flute Concerto G Major, II. Andante non Troppo
18	Sammartini - Sinfonia in A major J-C 63- II. Andante piano
19	Schubert - Symphony No. 9 in C major, D. 944- Andante-Allegro ma non troppo
20	Schubert -Trio D898 op.99 Andante
21	Vivaldi - Concerto in C major for 2 flutes, RV 533 - III Allegro
23	Brahms - Trio, piano, violon, and horn, mvt 2

Ai partecipanti è stato chiesto di raggruppare i brani ascoltati in gruppi a seconda della similitudine espressiva (sezione 1.3); è stato spiegato, in particolare, di porre attenzione alle intenzioni espressive percepite, all'esperienza emotiva provata durante l'ascolto: questo per evitare valutazioni che comportino un'organizzazione dei brani su base stilistica o su aspetti puramente strutturali; condizione, questa, più delicata se il partecipante ha una buona cultura musicale, come i soggetti catalogati come musicisti.

L'esperimento ha avuto una durata in media di 19 minuti, durante i quali i soggetti hanno potuto ascoltare i brani più di una volta, sia in fase di organizzazione e sia in fase di convalida dei gruppi predisposti. I soggetti non musicisti hanno avuto una durata maggiore ai musicisti, probabilmente causata dalla minor abitudine all'analisi di brani musicali, seppur a livello emotivo.

6.3 Elaborazione dati

Per l'elaborazione dei dati è stato adottato lo stesso procedimento presentato nel capitolo 5, che può essere suddiviso in quattro fasi principali:

- creazione di una matrice di dissimilarità in grado di relazionare i vari brani tra loro in base alle scelte di raggruppamento dei partecipanti;
- analisi di *scaling* multidimensionale per definire la spazialità ottimale a riprodurre con il minor numero possibile di dimensioni le distanze tra i vari oggetti;
- analisi di *bootstrapping* per confermare la bontà dei dati raccolti e la stabilità della soluzione trovata al punto precedente;
- analisi di raggruppamento per aggregare tra loro brani valutati similmente dal punto di vista delle intenzione espressive comunicate.

Infine sono state calcolate le caratteristiche più importanti riguardo al segnale audio ed è stata fatta un'attenta analisi per riconoscere correlazioni importanti tra le features calcolate e i risultati ottenuti dall'elaborazione dei dati raccolti.

6.3.1 Correlazione matrici dissimilarità

Dopo aver raccolto i dati relativi ai 40 partecipanti, è stata calcolata la matrice di distanze relativa alle dissimilarità tra i brani (per maggiori dettagli si rimanda alla sezione 5.3.1). La matrice quadrata ottenuta ha dimensione 22, e ogni elemento $[i, j]$ che la compone indica il numero di volte che il brano *i-esimo* è stato inserito in un gruppo differente da quello del brano *j-esimo;* i valori sulla diagonale maggiore non sono utili per le analisi successive che utilizzeranno questa matrice. I valori degli elementi della matrice di dissimilarità avranno quindi come massimo valore 20 (i due brani considerati sono sempre stati associati a gruppi diversi) e come valore minimo 0 (i brani sono risultati essere sempre nel

medesimo gruppo); il valore 20 è associato al numero di test effettuati per ogni categoria.

Sono state calcolate le matrici di dissimilarità separatamente per la categoria dei musicisti e per i non musicisti; questo è dovuto alla successiva analisi di correlazione tra le due matrici risultanti per evidenziare se vi sia una sostanziale omogeneità tra le scelte delle due categorie e se sia quindi possibile trattare i dati globalmente, senza perciò fare riferimento a due analisi separate.

È stato calcolato poi un indice della correlazione tra le due matrici. Per calcolare l'indice è stato utilizzato il test di Mantel che permette di calcolare la correlazione tra matrici di distanze. Sono stati eseguiti tre test di Mantel per il calcolo della correlazione tra la matrice di dissimilarità associata ai musicisti e la matrice dei non musicisti, e per ognuno è stato utilizzato una differente soluzione di calcolo della correlazione tra le coppie di variabili, ossia:

- *Pearson product-moment correlation coefficient*
- *Spearman's rank correlation coefficient*
- *Kendall rank correlation coefficient*

Questi indici di correlazione variano da un valore massimo di 1, interpretato come una totale correlazione tra le variabili x e y analizzate, fino a un valore minimo di -1 che significa correlazione negativa; un valore di indice pari a 0 è da interpretarsi come indipendenza tra le variabili.

I risultati dell'analisi di correlazione sono i seguenti: r=.78, r=.75, r=.65, rispettivamente. Il valore positivo e piuttosto elevato dell'indice di correlazione trovato indica che vi sia una relazione buona tra la valutazione data dai musicisti e le scelte dei non musicisti; per questo motivo è stato possibile calcolare una matrice totale di dissimilarità riguardante tutti i dati dei 40 partecipanti, senza distinzione di categoria.

Questo indice di correlazione elevato tra le matrici di dissimilarità indica inoltre una valutazione piuttosto simile e coerente dei brani da parte delle due categorie di partecipanti. Una correlazione prossima allo zero avrebbe implicato differenze notevoli di risultato tra i musicisti e i non musicisti, evidenziando una diversa percezione dell'intenzione espressiva musicale determinata da una diversa preparazione musicale.

6.3.2 *Multidimensional Scaling*

Data una matrice di dissimilarità dei dati è possibile applicare un'analisi di *scaling* multidimensionale (MDS) per identificare lo spazio a bassa dimensionalità che descrive in maniera migliore le relazioni tra le entità considerate; nello specifico le entità sono rappresentate dai 22 estratti musicali, e si deve considerare la soluzione dell'analisi ottimale in rispetto a ipotesi di bontà definite inizialmente.

La tecnica MDS usata è di tipo "non-metrico"; assume quindi che le misure di prossimità tra le entità analizzate e le distanze tra i punti sulla configurazione geometrica finale siano in relazione monotòna. In particolare è stata utilizzata,

anche per questa analisi, la tecnica di *scaling* multidimensionale di Kruskal (sezione 5.3.2), il quale propone un indice "di dissociazione", detto indice di *stress*, tra le prossimità iniziali e quelle della soluzione analitica, che gode anche della proprietà di invarianza rispetto a variazioni di scala degli assi ortogonali; questo indice è utilizzato inoltre anche dall'algoritmo di *scaling* per la valutazione della configurazione trovata, quando l'indice è accettabile si dice che la configurazione converge e l'analisi multidimensionale termina. L'indice di stress S (ricavato dalla formula 5.3) è interpretabile come varianza residua, esprime infatti quanta parte della variabilità delle prossimità osservate non è interpretata dalla configurazione ottenuta. Valori per S superiori a .20 sono da valutare come scarso adattamento dei dati alla dimensionalità proposta, e quindi solo configurazioni che diano indice di *stress* con valore minore sono accettabili.

Ogni analisi MDS è stata ripetuta più volte utilizzando un diverso numero di dimensioni: è necessario infatti inserire nei parametri di input la dimensionalità finale richiesta per lo *scaling*.

L'indice di *stress* di Kruskal relativo alla soluzione con una sola dimensione è stato 37%, e quindi non accettabile secondo quanto evidenziato precedentemente; la soluzione con 2 dimensioni ha portato un indice S del 15%, valore che rientra nella soglia di bontà proposta da Kruskal e la soluzione può essere accettata; per numero di dimensioni superiore le percentuali riscontrate sono state 12%, 9%, 7% rispettivamente per tre, quattro e cinque dimensioni.

La soluzione ottima dell'analisi MDS risulta essere la struttura bidimensionale, in quanto soluzioni con minor numero di dimensioni non sono accettabili e strutture a dimensionalità più elevata garantirebbero un incremento di adattabilità dei dati non importante, rispetto alla complessità che anche solo una dimensione aggiuntiva comporterebbe nella definizione della struttura stessa.

La configurazione risultante dall'analisi MDS totale per tutti i partecipanti è proposta in Figura 6.1. I brani sono rappresentati da punti e le etichette rispecchiano la disposizione degli estratti musicali proposta nella tabella 6.1. Le dimensioni sono etichettate con numeri per mantenere la neutralità nella loro definizione; successive analisi verificheranno le componenti più importanti nella comprensione dei due assi principali.

6.3.3 *Bootstrap Analysis*

Come visto nella sezione 5.3.3, una configurazione è stabile se non risente significativamente della presenza di errori accidentali di rilevazione e della casualità del campionamento delle unità statistiche.

In questa fase la configurazione ottenuta in fase di analisi MDS è stata testata seguendo la metodologia del *bootstrapping*, che ha l'obiettivo di verificare la coerenza dei dati raccolti. Per questa analisi, quindi, lo scopo è di verificare la presenza di eccessive differenze tra i raggruppamenti dei diversi partecipanti, le quali potrebbero influire negativamente sulla configurazione trovata dall'analisi di *scaling* multidimensionale; se l'analisi dovesse proporre valori poco simili e dispersivi, per i diversi brani, allora i dati raccolti sarebbero poco coerenti e non sarebbero statisticamente rilevanti.

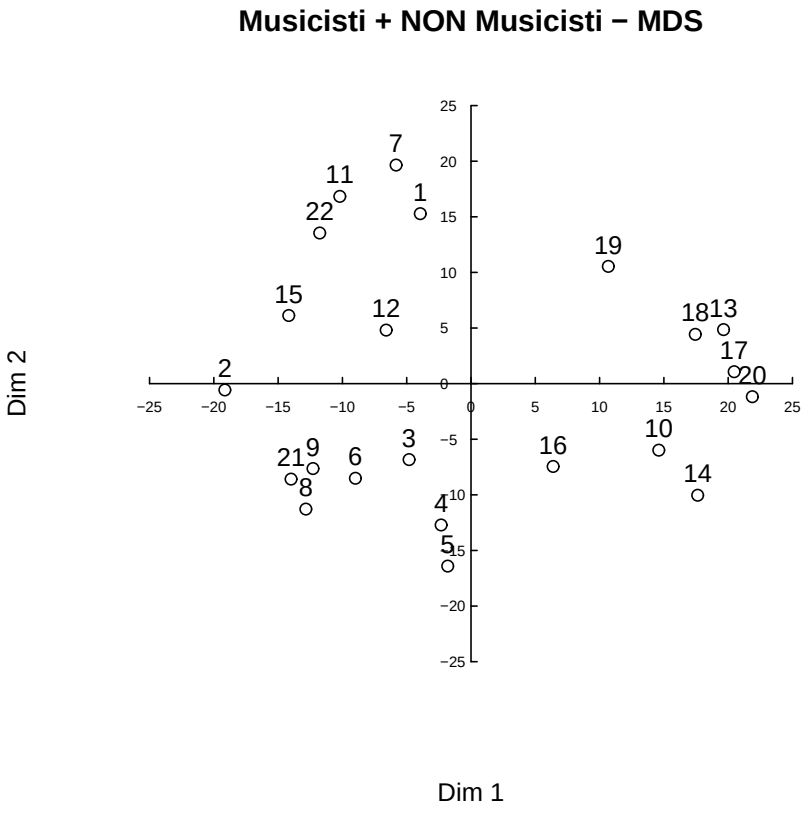


Figura 6.1: Analisi di *scaling* multidimensionale (MDS). Viene riportata la soluzione ottimale formata da una struttura bidimensionale; le distanze tra i brani rispecchiano le similitudini nel raggruppamento degli stessi derivanti dalle scelte dei partecipanti.

Tabella 6.2: Analisi di raggruppamento: i valori riportati sono relativi ai due metodi presentati in relazione al numero di cluster formati.

# cluster	2	3	4	5	6	7	8
PAM	.48	.53	.47	.44	.41	.42	.38
k-means	53 %	80 %	85 %	89 %	92 %	94 %	96 %

Sono state eseguite ciclicamente 200 iterazioni considerando di volta in volta un insieme di dati di partenza differente ricavato dal campionamento casuale, con reinserimento, dei dati iniziali (metodo Monte Carlo); l'insieme di dati per ogni iterazione è stato limitato a 40 valutazioni scelte casualmente. Inoltre a ogni insieme di dati è stata eseguita un'analisi di *scaling* multidimensionale con le stesse caratteristiche presentate nella sezione precedente.

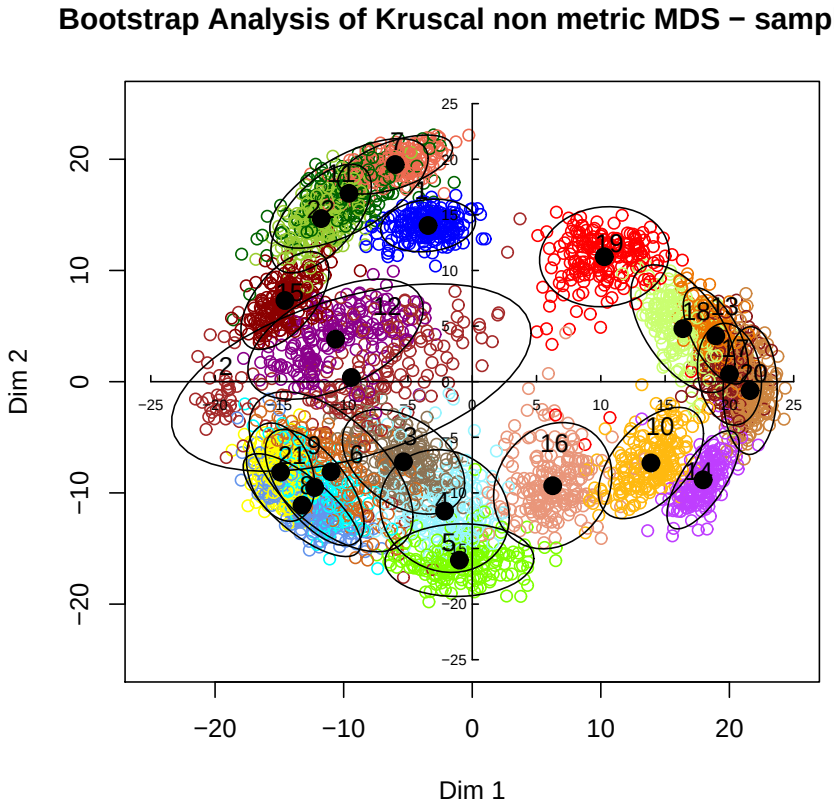
Per valutare la qualità della configurazione bidimensionale ricavata, relativamente alle scelte dei partecipanti, ne è stata quindi valutata la stabilità attraverso la sovrapposizione delle 200 configurazioni ricavate da altrettante iterazioni. Inoltre è stato calcolato per ogni brano un ellisse di confidenza che contenga la percentuale di valori desiderata; in Figura 6.2 viene presentato il risultato della sovrapposizione delle analisi con relativi ellissi di confidenza al 95% (grafico in alto); nella parte inferiore, della stessa figura, vengono proposti i soli ellissi senza i punti derivati dalle analisi MDS, in particolare in Figura 6.2(a) la percentuale di confidenza è al 95% mentre nel grafico a destra la percentuale è dell'85%.

Per ottenere questa rappresentazione è stata necessaria una rotazione degli assi in alcune iterazioni, conseguenza dell'analisi di *scaling* multidimensionale che può ruotare gli assi per una migliore disposizione dei risultati, fatto che impedirebbe la corretta visualizzazione di confronto per questo tipo di analisi.

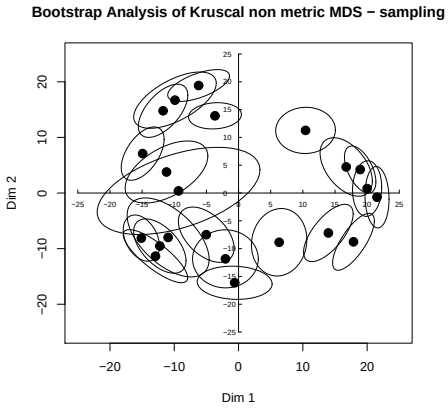
6.3.4 *Cluster Analysis*

Successivamente all'analisi di *scaling* multidimensionale è stata eseguita un'analisi di raggruppamento con lo scopo di creare gruppi con entità omogenee (*cluster analysis*). L'analisi di raggruppamento, come visto nella sezione 5.3.4 non richiede la forma delle relazioni tra le variabili in input, e quindi si adatta anche a un caso, come quello in esame, in cui non vi è conoscenza iniziale delle relazioni tra brani. In questa analisi sono proposte e messe a confronto due tipi di *cluster analysis*: l'analisi *k-means* e la PAM *clustering*.

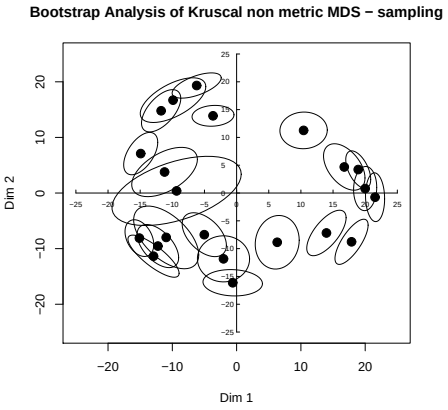
Il metodo di raggruppamento PAM è stato eseguito direttamente sulla matrice di dissimilarità, infatti non considera la configurazione dell'analisi MDS ma valuta le relazioni tra entità attraverso la matrice delle distanze; questa soluzione ha il vantaggio di definire la soluzione di raggruppamento migliore a prescindere dall'analisi di *scaling* multidimensionale che verrà fatta. In Tabella 6.2 sono elencati i risultati dell'analisi PAM: i valori indicano il grado medio di appartenenza di un oggetto a un *cluster*, quindi un valore più alto possibile è da preferirsi perché indica una soluzione migliore di raggruppamento. Si nota facilmente come la soluzione migliore sia rappresentata da 3 *cluster,* con una



(a) *Bootstrap analysis* relativa all'analisi MDS di Figura 6.1 per tutti i partecipanti con ellissi di confidenza al 95% per tutti i brani.



(b) Ellissi di confidenza al 95%.



(c) Ellissi di confidenza all' 85%.

Figura 6.2: *Bootstrap analysis* con ellissi di confidenza. Sono proposte le soluzioni per tutti i partecipanti con la visualizzazione dei valori delle 200 iterazioni (a), e con i soli ellissi di confidenza al 95% (b) e all'85% (c).

notevole differenza rispetto alle soluzioni con numero di raggruppamenti più elevato.

L'utilizzo della tecnica *k-means* comporta la scelta a priori del numero di *cluster* da formare; sono quindi state fatte più analisi di raggruppamento per stabilire quale soluzione fosse la più adatta allo spazio bidimensionale ricavato dall'analisi MDS. Sempre in Tabella 6.2 si vedono elencati i valori di *explained variance* rispetto a soluzioni che vanno da soli due gruppi formati a otto. È evidente come anche in questo caso la soluzione ottimale è la composizione di tre gruppi, infatti con un numero superiore di gruppi il valore aumenta ma di poche unità per ogni dimensione successiva, e il valore percentuale con due gruppi non è accettabile.

Infine è stato verificato che i gruppi formati dai due metodi siano gli stessi. In Figura 6.3a è possibile vedere la soluzione di raggruppamento comune a entrambi gli approcci. Inoltre è stata eseguita la stessa analisi anche per i dati separati per categorie, per scoprire eventuali brani valutati in maniera differente e/o configurazioni di raggruppamento diverse. In Figura 6.3b viene proposta la soluzione ottimale con 3 gruppi per i non musicisti e in Figura 6.3c la soluzione per i musicisti. Si nota da questi due ultimi grafici come il brano numero 12 venga inserito in un cluster diverso per le due categorie, fatto che verrà analizzato nella sezione 6.5.

6.4 Estrazione features

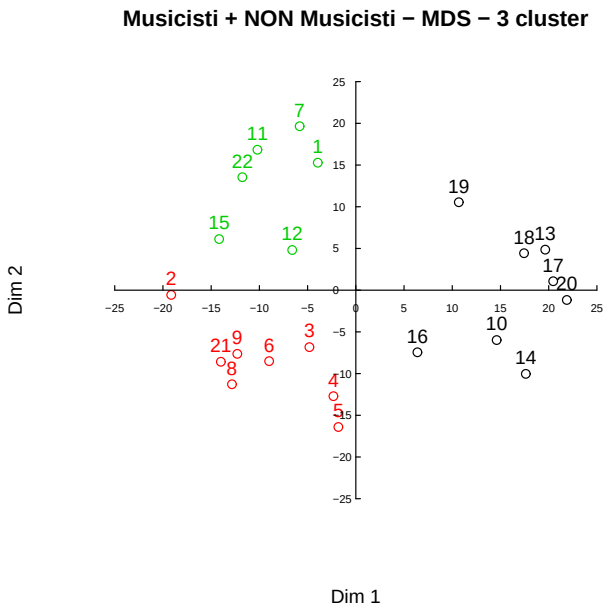
In questa fase sono state calcolate le caratteristiche relative al segnale audio per tutti gli estratti considerati nell'esperimento. Le features calcolate appartengono a un insieme di caratteristiche importanti per la discriminazione delle diverse intenzioni espressive (Sloboda e Juslin, 2001), e sono state usate in vari lavori presenti in letteratura sia per la classificazione dello stile musicale (Dannenberg et al., 1997) sia per la determinazione dei contenuti espressivi nelle performance musicali (Friberg et al., 2002; Mion e De Poli, 2008).

Le features calcolate sono sia di tipo locale e sia riguardanti interi eventi. Le features a livello locale sono calcolate usando frame senza overlap di 46 ms di lunghezza e poi sono stati considerati i valori medi valutati su finestre della durata di 4 s con overlap di 3.5 s. Le features calcolate sugli eventi hanno riguardato solo il valore derivante da finestre di 4 s con 3.5 s di sovrapposizione.

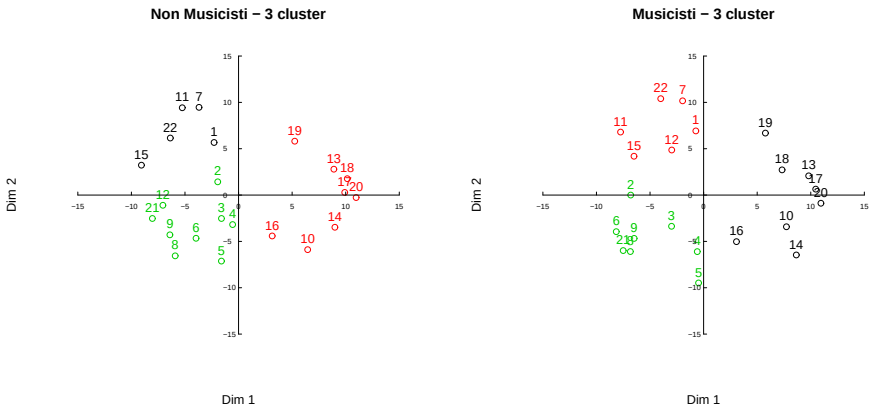
Alcune caratteristiche sono state utilizzate nell'esperimento illustrato nel capitolo 5, e sono: zerocross, RMS, roughness, spectral ratio (alto e medio), spectralflux, centroide, brightness, rolloff, onsets e attacco (sezione 5.4).

In aggiunta alle features già elencate ne sono state calcolate altre 3, con l'ipotesi che potessero essere importanti ai fini dell'analisi finale dei risultati e della comprensione della struttura ricavata.

1. **Lowenergy**. È interessante valutare come si distribuisce temporalmente l'energia del segnale per verificare se rimane costante per tutta la durata del brano se alcuni frame sono contrastanti con altri; un metodo per stimare questa statistica è calcolare l'indice di lowenergy, ossia la percentuale



(a) *Cluster analysis* relativa a tutti i patecipanti all'esperimento.



(b) *Cluster analysis* per i non musicisti.

(c) *Cluster analysis* per i musicisti.

Figura 6.3: Grafici relativi all'analisi di raggruppamento. La soluzione proposta, a seguito dell'analisi, vede la definizione di 3 cluster sia per la configurazione totale (a) sia per le due configurazioni relative alle diverse categorie, (b) non musicisti e (c) musicisti.

di frame che mostrano un comportamento energetico inferiore alla media (Tzanetakis e Cook, 2002).

2. **BeatSpectrum**. Questa features rappresenta una misura di similarità acustica in funzione dei vari intervalli su cui viene calcolata; brani molto strutturati o con passi ripetitivi avranno un alto valore di beat spectrum. Viene calcolato sul segnale audio in tre passi principali: si ricava la rappresentazione spettrale del segnale, ossia una sequenza di vettori; viene creata una matrice delle distanze per trovare le similitudini tra tutte le possibili coppie di vettori caratteristici; infine viene calcolato il beat spectrum analizzando le periodicità della matrice di similarità, attraverso l'autocorrelazione (Foote et al., 2002).

3. **EventDensity**. Stima la frequenza di eventi media del segnale audio, ossia il numero di note onsets per secondo. Soprattutto questa features sarà importante ai fini della comprensione di questo esperimento.

Sono stati utilizzati script Matlab per l'estrazione delle features, e in particolare le funzioni della toolbox MIR (Lartillot, 2010).

6.5 Discussione risultati

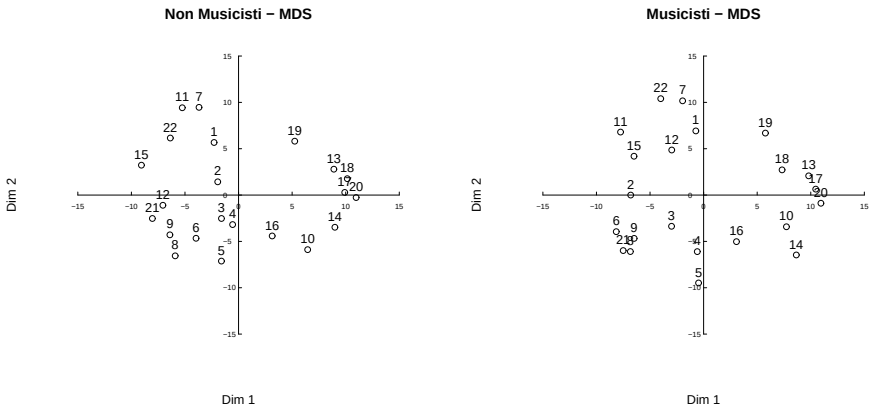
In questa sezione vengono discussi i risultati ottenuti dalle analisi effettuate e illustrate precedentemente in questo capitolo.

L'indice di correlazione calcolato per le matrici di dissimilarità delle due categorie di partecipanti (r=.78) indica una certa omogeneità nella valutazione dei brani da parte dei musicisti e dei non musicisti; l'elevata correlazione positiva indica coerenza nelle scelte organizzative fatte per l'esperimento. Come è stato evidenziato anche per l'esperimento presentato nel capitolo 5 e come viene trattato più avanti nella discussione, vi sono alcuni estratti che più di altri sono stati percepiti in maniera piuttosto diversa e quindi l'indice di correlazione è stato limitato in parte da questa difficoltà di comprensione.

Dopo aver verificato che i risultati appartenenti ai vari soggetti non siano influenzati dalla diversa preparazione in ambito musicale si sono potuti analizzare i dati raccolti nella loro totalità. L'analisi di *scaling* multidimensionale (Figura 6.1) ha evidenziato come soluzione ottimale una struttura bidimensionale. Una prima interpretazione di questo risultato, senza usare ulteriori strumenti di analisi, è che lo spazio tridimensionale fondamentale per la definizione delle intenzioni espressive della musica, dopo aver escluso le due features più importanti per la discriminazione dei brani, si riduce a due dimensioni che in parte sono il risultato delle variabili caratteristiche del terzo asse (sezione 2.2) e in parte sono dovute ai residui delle componenti minoritarie dei primi due assi (in particolare l'asse relativo all'attivazione).

Sono state proposte ulteriori analisi MDS separatamente per le due categorie (Figura 6.4).

In questo confronto si nota come la maggior parte dei brani risulti posizionata in una zona molto simile per entrambe le analisi MDS. Vi sono casi particolari



(a) MDS per la categoria dei non musicisti.

(b) MDS per la categoria dei musicisti.

Figura 6.4: MDS per le due categorie di partecipanti. Si nota come ci sia stabilità nella collocazione dei brani, con solo alcune eccezioni.

in cui questa similitudine non compare e vi è una differenza piuttosto marcata tra le posizioni dello stesso brano nelle due configurazioni; è il caso dei brani 2 e 12, i quali non presentano particolari features a livello di segnale audio che possano aver confuso la scelta dei soggetti; le differente valutazione può spiegarsi solo a causa di una diversa conoscenza musicale dei partecipanti, che hanno interpretato in maniera differente le intenzioni espressive di questi due brani.

L'analisi di *bootstrapping* ha validato la soluzione ricavata dallo *scaling* MDS. Come riportato in Figura 6.2, il risultato delle 200 iterazioni con sovrapposizione ha portato ad avere configurazioni piuttosto simili per quasi tutti i brani. La bontà di questa analisi è stata valutata con un metodo qualitativo, l'ellisse di confidenza, ripostato sia per una percentuale del 95% e sia per un valore di 85%. Si notano però alcuni brani con valori particolarmente dispersivi e conseguente ellisse di confidenza molto esteso e sono i numeri 2 e 12; in particolare la dispersione ricavata per il primo è piuttosto accentuata. È stata eseguita un'ulteriore analisi di *bootstrapping*, meno casuale della prima, che mette in luce alcuni risultati appena esposti: l'analisi si è ripetuta per 200 iterazioni, come la prima, però l'insieme di dati di partenza è stato di volta in volta lo stesso dei valori iniziali a meno dell'apporto della valutazione di un partecipante, estratto casualmente (*one-out bootstrap analysis*); in Figura 6.5 si nota più chiaramente la stabilità delle posizioni dei brani, con esclusione dei già citati brani numero 2 e 12, i quali mostrano un comportamento dispersivo nonostante la minor casualità nella selezione dei dati per questa analisi.

Ai fini della comprensione delle caratteristiche dei due assi principali, quindi, è stata posta attenzione nel valutare separatamente il contributo dei brani 2 e

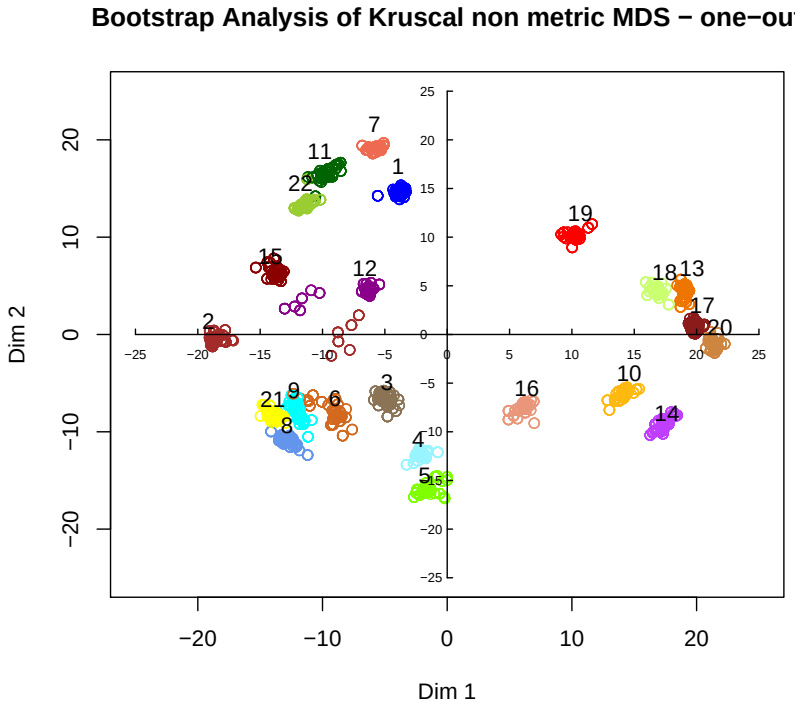


Figura 6.5: *Bootstrap analysis* con esclusione di una valutazione per ogni iterazione (200 cicli). Viene evidenziata la stabilità della configurazione e l'eccezione presentata dai brani 2 e 12.

Tabella 6.3: Correlazione calcolata tra le due dimensioni risultanti dall'analisi MDS e le features acustiche estratte dai segnali audio dei brani considerati.

#	Zerocross	Lowenergy	RMS	Roughness	SRh	SRm	SF
dim1	-.30	.16	-.51	-.54	-.09	-.01	-.55
dim2	.17	-.23	.20	.30	.07	-.17	.15
	Centroid	Brightness	Rolloff	Onsets	Attacks	BS	ED
dim1	-.36	-.32	-.41	.16	.37	.40	-.56
dim2	.02	.20	.04	-.15	.16	.31	.19

12 che sono stati poco coerentemente valutati dai partecipanti. Per gli altri brani invece l'omogeneità riscontrata è un fattore importante per indagare le caratteristiche che hanno comportato questa organizzazione collettiva a livello mentale.

L'analisi di raggruppamento in questo caso ha proposto tre gruppi principali (Figura 6.3a); anche il calcolo separato per i dati relativi ai musicisti e ai non musicisti ha portato alla definizione degli stessi *cluster*, con la sola eccezione del brano 12 che viene associato a due gruppi diversi nelle due configurazioni; quest'ultimo aspetto è legato all'analisi MDS fatta in precedenza per le due categorie.

L'analisi della correlazione tra le features calcolate per i brani e i gruppi rilevati dalla *cluster analysis* non ha un valore significativo come può invece essere per la relazione che intercorre tra le features e le due dimensioni principali della struttura bidimensionale; i valori degli indici ricavati sono presentati in Tabella 6.3. È da notare come non vi siano features correlate in modo dominante ad una delle due dimensioni; in particolare la dimensione 2 non presenta valori di correlazione significativi e quindi non è interpretabile con questa analisi ma verrà valutata attraverso ascolti mirati alla comprensione delle caratteristiche dei brani rispetto a tale asse. La correlazione calcolata in modo diretto tra una feature e una dimensione assume valori compresi tra -1 e 1, come visto in altre sezioni; per questa correlazione un valore compreso tra .3 e .7 viene definito come correlazione moderata tra le due variabili; non vi sono valori di forte correlazione (r>.7), vi sono invece alcune relazioni importanti tra la prima dimensione e le features relative a: RMS, roughness, spectralflux ed eventdensity (r=-.51, r=-.54, r=-.55, r=-.56 rispettivamente). Queste caratteristiche del segnale definiscono un criterio di scelta nell'organizzazione dei brani da parte dei partecipanti, in particolare sono da valutare gli aspetti legati alla correlazione negativa con roughness ed eventdensity (NPS: *Note Per Second*). La roughness esprime una misura di dissonanza: se due toni musicali suonano insieme i disturbi nell'armonia sono in parte causati dai battiti dei loro toni parziali, che si combinano generando un suono più ruvido; questa feature, correlata alla dimensione 1, determina una selezione dei brani in base alla loro dissonanza. La eventdensity, che sarà chiamata in seguito semplicemente NPS, in prima approssimazione esprime la rapidità di esecuzione del brano; in questo esperimento in cui sono stati usati brani con lo stesso metronomo NPS è interpretabile come

"metronomo percepito". In questo senso la correlazione di questa feature con la prima dimensione indica la classificazione da parte dei soggetti in base a un metronomo inverso percepito. Quest'ultima osservazione suggerisce che la feature NPS sia la componente principale per la determinazione del fattore di attivazione nella struttura generale ottenuta in mancanza del fattore fondamentale che è il tempo.

La prima dimensione, così analizzata, propone parte dei contenuti della dimensione di *activity* rimasta in seguito all'esclusione del tempo, e parte descritta da una maggiore "armonicità" dei brani musicali, oltre che a features legate a una maggiore energia e a cambiamenti locali relativi allo spettro (RMS e spectralflux).

Infine, sono state effettuate prove di ascolto dei brani, da parte dell'autore di questo lavoro e da esperti, secondo le due dimensioni; queste prove hanno confermato le osservazioni derivanti dall'analisi fatta per quanto riguarda l'asse della prima dimensione, e hanno inoltre portato all'individuazione di una descrizione per la seconda dimensione specificata dalla diversa strumentazione dei brani: valori positivi del secondo asse vengono associati a estratti orchestrali, mentre valori negativi sono relativi a solisti o con una composizione strumentale ridotta.

Capitolo 7

Conclusioni

I due esperimenti proposti in questa tesi hanno portato una serie di risultati molto interessanti per lo studio dettagliato di una struttura a bassa dimensionalità per descrivere l'organizzazione delle emozioni nella musica percepite da un generico ascoltatore.

La prima considerazione è relativa alla percezione emotiva che categorie di persone con differente cultura musicale sperimentano. In entrambi gli esperimenti, infatti, l'elevato indice di correlazione tra le scelte dei musicisti e dei non musicisti implica una valutazione dei brani omogenea per entrambi. Risultato, questo, che conferma le analisi di Bigand (2005).

Relativamente al primo esperimento, che considera solo brani in modalità maggiore, è stato definito uno spazio a due dimensioni: importante sottolineare come la prima dimensione sia fortemente correlata al metronomo inverso, riscontrando così un risultato accettato in molti lavori presenti in letteratura, dove viene definita questa dimensione principalmente con il termine *attivazione*. La seconda dimensione identificata non ha presentato valori di correlazione significativi rispetto alle feature acustiche estratte (Tabella 7.1a).

Il secondo esperimento ha messo in luce le caratteristiche di riferimento per l'organizzazione delle emozioni della musica in mancanza delle due features principali. Questo ha permesso l'analisi delle componenti residue derivanti dai due assi principali, *valence* e *activity*, e la comprensione delle features che strutturano la terza dimensione, molto discussa in letteratura (sezione 2.2). È da sottolineare come la prima dimensione ottenuta sia correlata con la densità di eventi (NPS) del segnale audio e con la roughness (Tabella 7.1b):

- la feature NPS correlata alla prima dimensione indica come un soggetto generico durante l'ascolto percepisca una diversa attivazione nonostante la privazione della componente principale per questo scopo, che è il Tempo; rappresenta la componente più importante, dopo il metronomo, per la definizione della dimensione relativa all'a*rousal* (Bigand et al., 2005) e il fattore di cinetica residuo per la *kinetics* (Canazza et al., 2003b);

Tabella 7.1: Indici qualitativi di correlazione tra le dimensioni ottenute e le features acustiche estratte dagli stimoli musicali per: (a) esperimento 1 (cfr sezione 5.5) e (b) esperimento 2 (cfr sezione 6.5).

(a) Esperimento 1.

#	Tempo	Brightness	Rolloff	Zerocross
dim 1	- - -	- - -	-	- -
dim 2	-	++	++	+

(b) Esperimento 2.

#	RMS	Roughness	Spectralflux	NPS
dim 1	- -	- -	- -	- -
dim 2	+	+	+	+

- la roughness, che rappresenta la dissonanza calcolata relativamente ad un estratto musicale, è stata associata molte volte in letteratura all'espressione della tensione musicale (Bigand et al., 1996; Vassilakis e Kendall, 2008; Pressnitzer et al., 2000); i risultati di questo secondo esperimento confermano la teoria di una terza dimensione associata alla *tension*, meno definita ma di fondamentale importanza nella discriminazione delle intenzioni espressive, come visto in sezione 2.2.

La seconda dimensione, risultante dalle analisi del secondo esperimento, è fortemente relazionata con la strumentazione dei brani, e non è stata rilevata alcuna correlazione significativa con le features acustiche estratte. Questo risultato implica la presenza di ulteriori dimensioni minoritarie per la struttura delle emozioni, poco correlate a features del segnale audio, comunemente usate in questo tipo di analisi, ma piuttosto interpretate da aggettivi che rispecchiano l'idea dell'ascoltatore di come sia eseguito il brano, per esempio differenziando un'esecuzione orchestrale da un assolo.

Gli esperimenti proposti seguono un approccio di analisi *score-dependent*. Uno sviluppo futuro per queste analisi è senz'altro la possibilità di individuare e definire uno spazio di astrazione intermedia che sia valido per la discriminazione delle intenzioni espressive sia per analisi relative alla partitura sia per analisi basate sulla performance dell'esecutore (*score-independent*).

I domini applicativi per questi risultati sono diverse, tra cui le più importanti: (a) la possibilità di reperire contenuti musicali con ricerche basate sulle emozioni dell'utente, portando alla creazione di algoritmi ottimizzati che riescano a superare le barriere di ricerca finora incontrate, che limitano ad attributi basati su etichette definite dal personale; prodotti come Genius di iTunes[1] avrebbero un aumento di fidelizzazione da parte dell'utenza; (b) RENCON, definizione di modelli computazionali per la performance automatica.

[1] http://www.apple.com/itunes/

Bibliografia

Bigand, E., Parncutt, R., e Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, 58(1):125–141. 7

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., e Dacquet, A. (2005). Multidimensional Scaling of Emotional Responses to Music: The Effect of Musical Expertise and of the Duration of the Excerpts. *Cognition and Emotion*, 19(8):1113–1139. 4, 4.1, 4.1, 4.1.3, 4.2.2.1, 5.2, 5.2.2, 5.3, 5.5.1, 6.2, 6.2.2, 7

Bresin, R. e Friberg, A. (2010). Emotion rendering in music: range and characteristic values of seven musical variables. *Cortex (accepted for publication)*. 3.1.3

Camurri, A., De Poli, G., Leman, M., e Volpe, G. (2005). Communicating Expressiveness and Affect in Multimodal Interactive Systems. *IEEE Multimedia*, 12(1):43–53. 3.1.3

Canazza, S., De Poli, G., Drioli, C., Rodà, A., e Vidolin, A. (2000). Audio morphing different expressive intentions for Multimedia Systems. *IEEE Multimedia*, 7(3):79–83. 3.2.3

Canazza, S., De Poli, G., Drioli, C., Roda', A., e Vidolin, A. (2004). Modeling and Control of Expressiveness in Music Performance. *IEEE Multimedia*, 7(3):79–83. 3.1.3

Canazza, S., De Poli, G., Mion, L., Roda', A., Vidolin, A., e Zanon, P. (2003a). Expressive Classifiers at CSC: An overview of the main research streams. *In Proc. of the XIV Colloquium on Musical Informatics (XIV CIM 2003)*, pages 64–68. 3.1.3

Canazza, S., De Poli, G., Rinaldin, S., e Vidolin, A. (1997). Sonological analysis of clarinet expressivity. In Leman, M., editor, *Music, Gestalt, and Computing - Studies in Cognitive and Systematic Musicology*, pages 431–440. Berlin, Heidelberg: Springer-Verlag. 3.2.3

Canazza, S., De Poli, G., Rodà, A., e Vidolin, A. (2003b). An Abstract Control Space for Communication of Sensory Expressive Intentions in Music Performance. *Journal of New Music Research*, 32(3):281–294. 3.2.3, 3.3, 3.4, 3.5, 4.2.2.1, 7

Canazza, S., Rodà, A., e De Poli, G. (2010). On the espressive gestures: looking for common traits between musical and physical domain. *In Proc. of Kansei Engineering and Emotion Research*, pages 1589–1597. 4.2

Cox, T. e Mackay, C. (1985). The measurement of self-reported stress and arousal. *British Journal of Psychology*, 76:183–186. 2.2.4

Damasio, A. R. (2003). *Looking for Spinoza: Joy, sorrow and the feeling brain.* New York: Harcourt. 4.1.2.2

Dannenberg, R., Thom, B., e Watson, D. (1997). A machine learning approach to musical style recognition. *In proceedings of the International Computer Music Conference (ICMC'97)*, pages 344–347. 5.4, 6.4

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals.* John Murray. 1.1

De Poli, G. (2004). Methodologiesfor expressiveness modelling of and for music performance. *Journal of New Music Research*, 33(3):189–202. 4.2

Fenza, D., Mion, L., Canazza, S., e Roda', A. (2005). Physical movement and Musical gesture: A multilevel mapping strategy. *In CDROM Proc. of Sound and Music Computing Int. Conf. (XV CIM, SMC05)*. 3.1.3

Foote, J., Cooper, M., e Nam, U. (2002). Audio Retrieval by Rhythmic Similarity. In *ISMIR 2002*. 2

Friberg, A., Schoonderwaldt, E., Juslin, P., e Bresin, R. (2002). Automatic real-time extraction of musical expression. *In proceedings of the International Computer Music Conference (ICMC'02)*, pages 365–367. 5.4, 6.4

Gabrielsson, A. (2001). Emotions in strong experiences with music. In Sloboda, J. A. e Juslin, P. N., editors, *Music and emotion: theory and research*, pages 431–449. New York: Oxford University Press. 4.1.1

Gaver, W. W. e Mandler, G. (1987). Play it again, Sam: On linking music. In *Cognition and Emotion*, volume 1, pages 259–282. 1.1.3

Hashimoto, S. (1997). KANSEI as the third target of information processing and related topics in Japan. *In Proc. of Intl. Workshop on Kansei - Technology of emotion*, pages 101–104. 3.1.2

Imberty, M. (1986). *Suoni Emozioni Significati.* Bologna: CLUEB. 1, 1.1.3

James, W. (1884). What is an Emotion? *Mind*, 9:188–205. 1.1

Konecni, V. J. (1979). Determinants of aesthetic preference and effects of exposure to aesthetic stimuli: Social, emotional, and cognitive factors. *Progress in experimental personality research*, 9:149–197. 1.1.3

Kruskal, J. B. (1964). Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29:1–27. 5.3.2

Laban, R. e Lawrence, F. C. (1947). *Effort.* London: Macdonald & Evans Ltd. 3.1.3

Langner, J. e Goebl, W. (2003). Visualizing Expressive Performance in Tempo-Loudness Space. *Computer Music Journal*, 27(4):69–83. 3.2.3

Larsen, R. J. e Diener, E. (1992). Promises and problems with the circumplex model of emotion. In Clark, M. S., editor, *Review of personality and social psychology: Emotion*, volume 13, pages 25–59. Newbury Park, CA: Sage. 2.1.1.2

Lartillot, O. (2010). *MIRtoolbox 1.3 - User's Manual.* Finnish Centre of Excellence in Interdisciplinary Music Research. 6.4

Leman, M. (2000). Visualization and calculation of roughness of acoustical musical signals using the synchronization index model (SIM). *In Proc. COST G-6 Conf. Digital Audio Effects (DAFX-00)*, pages 125–130. 2

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297. 5.3.4

MacRae, A. W., Howgate, P., e Geelhoed, E. (1990). Assessing the similarity of odours by sorting and by triadic comparison. *Chemical Sense*, 15:691–699. 1.3

Matthews, G., Jones, D. M., e Chamberlain, A. G. (1990). Refining the measurement of mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology*, 81:17–42. 2.2.1, 2.2.4, 2.4

Mion, L. e De Poli, G. (2008). Score-independent audio features for description of music expression. *IEEE Trans. Speech, Audio and Language Process*, 16(2):458–466. 4.2.1, 4.2, 4.2.2.2, 5.4, 6.4

Mion, L., De Poli, G., e Rapanà, E. (2010). Perceptual organization of affective and sensorial expressive intentions in music performance. *ACM Transactions on Applied Perception (TAP)*, 7(2). 4, 4.2, 4.2.2, 4.3, 4.4

Oatley, K. (1992). *Best laid schemes. The psychology of emotion.* Cambridge, MA: Harvard University Press. 1.1.2

Paganin, N., Benacchio, A., Locascio, F., e Cassano, G. (2010). Rapporto interno CSC Padova. Technical report, Università degli Studi di Padova. 5.1

Picard, R. (1997). *Affective Computing.* Cambridge, MA: the MIT Press. 3.1.1

Plutchik, R. (1994). *The psychology and biology of emotion.* New York: Harper Collins College Publisher. 1.1.2

Pressnitzer, D., McAdams, S., Winsberg, S., e Fineberg, J. (2000). Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception & Psychophysics*, 62(1):66–80. 7

Ramirez, R., Maestre, E., e Serra, X. (2011). A Rule-Based Approach to Music Performance Modelling. *IEEE Transactions on Evolutionary Computation (in press)*. 3.1.3

Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotion. *Journal of Personality and Social Psychology*, 67:525–539. 2.2.1

Revelle, W. e Rocklin, T. (1979). Very Simple Structure: an alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioural Research*, 14:403–414. 2.2.4

Rosch, R. (1978). Principles of categorization. In *Cognition and categorization*, pages 27–48. Hillsdale, NJ: Erlbaum. 1.1.2

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178. 2, 2.1.1.1, 2.2.1, 4.2.2.1

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110:145–172. 2.1.1, 2.1.1.1

Russell, J. A. e Fehr, B. (1984). Concept of Emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, 113(3):464–486. 1.1

Russell, J. A. e Feldman Barrett, L. (1999). Core Affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, 76:805–819. 2.1, 2.1, 2.2.1

Scherer, K. R. (1994). Affect bursts. In van Goozen, S., van de Poll, N. E., e Sergeant, J. A., editors, *Emotions: Essays on emotion theory*, pages 161–196. Hillsdale, NJ: Erlbaum. 4.1

Schimmack, U. e Grob, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality*, 14:325–345. 2.2.1, 2.2

Schimmack, U. e Reinsenzein, R. (2002). Experiencing Activation: Energetic Arousal and Tense Arousal are not mixture of Valence and Activation. *Emotion*, 2(4):412–417. 2.2.2, 2.3

Serra, X. (1997). Musical sound modeling with sinusoids plus noise. In Roads, C., Pope, S. T., Piccialli, A., e De Poli, G., editors, *Musical Signal Processing*, pages 91–122. Swets & Zeitlinger, The Netherlands. 4.2.2.2

Sjöberg, L., Svensson, E., e Persson, L. O. (1979). The measurement of mood. *Scandinavian Journal of Psychology*, 20:1–18. 2.2.1, 2.2.3, 2.2.4

Sloboda, J. A. e Juslin, P. N. (2001). *Music and Emotion: Theory and Research.* New York: Oxford University Press. 1.1, 4.2.2.2, 5.4, 6.4

Snyder, B. (2000). *Music and Memory.* Cambridge: The MIT Press. 3

Steyer, R., Schwenkmezger, P., Notz, P., e Eid, M. (1994). Theoretical analysis of a multidimensional mood questionnaire (MDBF). *Diagnostica*, 40:320–328. 2.2.1

Suzuki, K., Camurri, A., Ferrentino, P., e Hashimoto, S. (1998). Intelligent Agent System for Human-Robot Interaction through Artificial Emotion. *Proc. IEEE Intl. Conf. On Systems Man and Cybernetics SMC'98.* 3.1.2

Suzuki, K., Hikiji, R., e Hashimoto, S. (2002). Development of an Autonomous Humanoid Robot, iSHA, for Harmonized Human-Machine Environment. *Journal of Robotics and Mechatronics*, 14(5):324–332. 3.1.2

Thayer, R. E. (1989). *The biopsychology of mood and activation.* New York: Oxford University Press. 2, 2.1.1.2, 2.1.2, 2.2.1, 2.2.1, 2.2.2, 2.2.4

Tzanetakis, G. e Cook, P. (2002). Musical genre classification of audio signals. *IEEE Tr. Speech and Audio Processing*, 10(5):293–302. 1

Varela, F., Thompson, E., e Rosch, E. (1991). *The embodied mind.* Cambridge, MA: MIT Press. 4.2

Vassilakis, P. (2001). Auditory roughness estimation of complex spectra - Roughness degrees and dissonance ratings of harmonic intervals revisited. *Journal of Acoustical Society of America*, 110. 5.4

Vassilakis, P. e Kendall, R. A. (2008). Auditory roughness profiles and musical tension-release patterns in a Bosnian ganga song. *Journal of the Acoustical Society of America*, 124:2448–2448. 7

Volpe, G. (2003). *Computational models of expressive gesture in multimedia systems.* PhD thesis, University of Padua. 3.1, 3.2, 3.1, 3.2

Watson, D. e Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98:219–235. 2.1.1.2, 2.1.2.2, 2.2.1

Watson, D., Wiese, D., Vaidya, J., e Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, 76:820–838. 2.2.1

Zwick, W. R. e Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99:432–442. 2.2.4