

**ALMA MATER STUDIORUM – UNIVERSITA' DI BOLOGNA
SEDE DI CESENA**

SECONDA FACOLTA' DI INGEGNERIA CON SEDE A CESENA
CORSO DI LAUREA IN INGEGNERIA INFORMATICA

**CLASSIFICAZIONE DI BRANI MUSICALI
MEDIANTE RETI NEURONALI ARTIFICIALI**

Elaborato in :
Fondamenti di Informatica B

Relatore:
Andrea Roli

.....

Presentato da:
Alberto Bernardini

.....

Sessione III

Anno Accademico 2011/2012

Indice

<u>Introduzione</u>	2
 PARTE PRIMA	
<u>1.0 Reti Neurali</u>	4
1.1 Reti Neurali Biologiche.....	5
1.2 Reti Neurali Artificiali.....	6
1.3 La Rete Feedforward.....	10
1.4 L'algoritmo di Backpropagation.....	13
1.5 Intelligenza Artificiale e Informatica Musicale : Simbolismo e Connesionismo a confronto.....	16
 <u>2.0 Percezione Sonora e Musica</u>	 19
2.1 Percezione Uditiva Umana.....	19
2.2 Note e Timbro.....	22
2.3 Armonia, Ritmo e Melodia.....	24
2.4 Illusioni Sonore.....	26
2.5 Differenze nel Riconoscimento di Pattern Musicali tra Uomo e Macchina.....	29
 <u>3.0 Classificazione di File Audio Mediante Reti Neurali</u>	 31
3.1 Classificazione Audio ed Intelligenza Artificiale.....	31
3.2 Criteri di Classificazione.....	33
3.3 Metodi di Classificazione: Formati Audio ed Estrazione di Feature.....	35
 <u>4.0 MIDI</u>	 41
4.1 Introduzione al MIDI e alla Rappresentazione Simbolica dell'Informazione Musicale.....	41
4.2 Interfaccia e Protocollo MIDI.....	43
4.3 Conformazione dei File MIDI.....	47

PARTE SECONDA

5.0 <u>Software Utilizzati per gli Esperimenti e Procedura di Esecuzione</u>	51
5.1 JSymbolic.....	51
5.2 JFeatureXmlExtractor.....	57
5.3 Fogli di Calcolo Costruiti per la Normalizzazione.....	58
5.4 JPatConverter.....	60
5.5 JavaNNS.....	63
6.0 <u>Esperimenti</u>	65
6.1 Bach e Mozart.....	65
6.2 Chet Baker e Luis Armstrong.....	78
6.3 Beatles e Rolling Stones.....	101
6.5 Beatles e Led Zeppelin.....	112
<u>Conclusione e Sviluppi Futuri</u>	136

Introduzione

L'informatica musicale è una disciplina in continua crescita che sta ottenendo risultati davvero interessanti con l'impiego di sistemi artificiali intelligenti, come le reti neurali, che permettono di emulare capacità umane di ascolto e di esecuzione musicale.

Di particolare interesse è l'ambito della codifica di informazioni musicali tramite formati simbolici, come il MIDI, che permette un'analisi di alto livello dei brani musicali e consente la realizzazione di applicazioni sorprendentemente innovative. Una delle più fruttifere applicazioni di questi nuovi strumenti di codifica riguarda la classificazione di file audio musicali.

Questo elaborato si propone di esporre i fondamentali aspetti teorici che concernono la classificazione di brani musicali tramite reti neurali artificiali e descrivere alcuni esperimenti di classificazione di file MIDI.

La prima parte fornisce alcune conoscenze di base che permettono di leggere gli esperimenti presenti nella seconda sezione con una consapevolezza teorica più profonda.

Il fine principale della prima parte è quello di sviluppare una **comparazione da diversi punti di vista disciplinari tra le capacità di classificazione musicale umane e quelle artificiali.**

Si descrivono le reti neurali artificiali come sistemi intelligenti ispirati alla struttura delle reti neurali biologiche, soffermandosi in particolare sulla rete Feedforward e sull'algoritmo di Backpropagation.

Si esplora il concetto di percezione nell'ambito della psicologia cognitiva con maggiore attenzione alla percezione uditiva. Accennate le basi della psicoacustica, si passa ad una descrizione delle componenti strutturali prima del suono e poi della musica: la frequenza e l'ampiezza delle onde, le note e il timbro, l'armonia, la melodia ed il ritmo. Si parla anche delle illusioni sonore e della rielaborazione delle informazioni audio da parte del cervello umano.

Si descrive poi l'ambito che interessa questa tesi da vicino: il MIR (*Music Information Retrieval*). Si analizzano i campi disciplinari a cui questa ricerca può portare vantaggi, ossia quelli commerciali, in cui i database musicali svolgono ruoli importanti, e quelli più speculativi ed accademici che studiano i comportamenti di sistemi intelligenti artificiali e biologici. Si descrivono i diversi metodi di classificazione musicale catalogabili in base al tipo di formato dei file audio in questione e al tipo di feature che si vogliono estrarre dai file stessi.

Conclude la prima sezione di stampo teorico un capitolo dedicato al MIDI che racconta la storia del protocollo e ne descrive le istruzioni fondamentali nonché la struttura dei midifile.

La seconda parte ha come obiettivo quello di descrivere gli esperimenti svolti che classificano file MIDI tramite reti neurali mostrando nel dettaglio i risultati ottenuti e le difficoltà incontrate.

Si coniuga una presentazione dei programmi utilizzati e degli eseguibili di interfaccia implementati con una descrizione generale della procedura degli esperimenti.

L'obiettivo comune di tutte le prove è **l'addestramento di una rete neurale in modo che raggiunga il più alto livello possibile di apprendimento circa il riconoscimento di uno dei due compositori dei brani che le sono stati forniti come esempi.**

PARTE PRIMA

1.0

Reti Neuronali

Il computer è estremamente più potente ed efficace dell'uomo per quel che concerne attività dall'iter predefinito quali il calcolo, la copia di elementi o la ricerca di dati in grandi database. Tuttavia l'uomo è in grado di svolgere azioni molto più complesse, dipendenti da un numero elevato di fattori in perenne mutazione, ottenendo risultati sorprendenti. Egli è capace di apprendere tramite esempi e confronta attivamente diverse esperienze passate con le difficoltà del presente, riadattando molto flessibilmente le conoscenze pregresse a nuovi problemi mai risolti [001].

La constatazione di questi deficit dei sistemi artificiali, verso la fine degli anni 1940, ha suggerito lo studio di nuove metodologie di computazione, ispirate ai criteri di deduzione e ragionamento delle reti neurali umane o animali e che ne imitano, in diversa misura, il funzionamento.

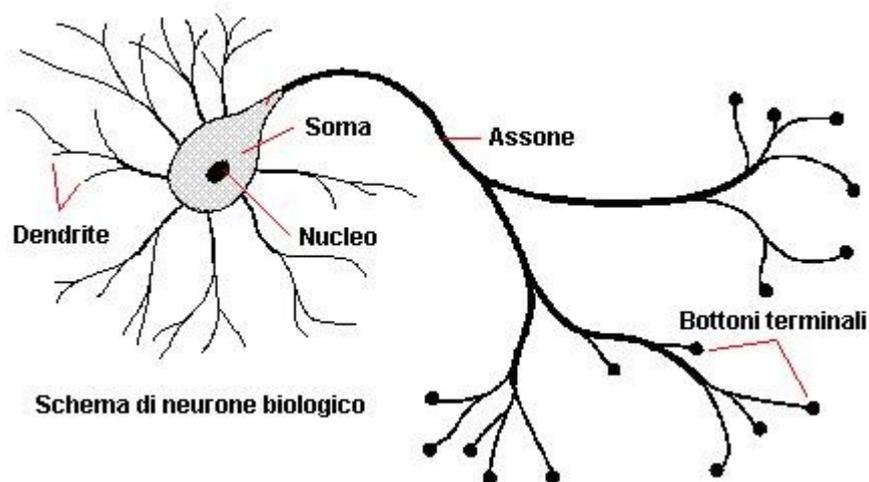
Questo capitolo descrive le reti neurali del cervello, per poi passare ad analizzare le reti neurali artificiali. In seguito si sofferma ad approfondire un particolare tipo di rete, quella Feedforward con l'algoritmo di apprendimento Backpropagation che è quella impiegata negli esperimenti di questa tesi. Infine presenta una panoramica delle correnti di pensiero nell'ambito delle scienze cognitive, in particolare il

simbolismo ed il connessionismo, sia in termini generali sia per quanto riguarda più propriamente la computer music. Le reti neuronali sono le protagoniste del connessionismo. Si descrivono i pregi ed i difetti di un loro utilizzo in applicazioni musicali.

1.1

Reti Neuronali Biologiche

Gran parte degli esseri viventi pluricellulari è dotata di cellule nervose (neuroni) capaci di recepire tramite i sensi stimoli esterni ed elaborarli al fine di apprendere o memorizzare [002]. Il cervello umano è sicuramente l'esempio più sorprendente. E' un intrecciato sistema di 1300 cmq circa che vanta più di 100 miliardi di neuroni ed un numero ancora maggiore di connessioni tra di essi dette sinapsi [003]. I neuroni e le sinapsi formano la cosiddetta rete neurale in cui viaggiano segnali elettrici a velocità nell'ordine dei 100 m/s [004].



I neuroni a loro volta sono suddivisibili in diverse parti:

-il **soma**: è il corpo cellulare costituito dal **pirenoforo** in cui risiede il **nucleo** e gli

altri organelli cellulari quali l'apparato del Golgi [005];

-l'**assone**: è una linea conduttrice di impulsi in direzione di uscita dal soma che nella sua porzione prossimale rimane unico, per poi diramarsi nella sua porzione distale [006];

-i **dendriti**: sono le fibre minori che si diramano dal neurone e trasportano i segnali nervosi dall'esterno verso il soma, per ogni neurone insieme formano “l'albero dendritico”. [007]

I neuroni comunicano tra loro poiché i loro assoni sono collegati ai dendriti degli altri tramite le **sinapsi**. Si forma così una intricatissima e complessa struttura a strati in cui i neuroni comunicano tra loro tramite segnali elettrici e chimici trasportati da assoni e dendriti.

Ogni neurone al suo interno compie una “somma pesata” dei segnali in ingresso. Se la somma supera una certa soglia il neurone si attiva e produce un “potenziale di azione” che è trasportato dall'assone. [008] La capacità di apprendimento della rete neurale è dovuta ad un cambiamento nel lungo periodo della forza delle connessioni. In riferimento a ciò le sinapsi si distinguono in due categorie: eccitatorie, ossia che innalzano il potenziale del neurone ed inibitorie che, invece, lo abbassano. [009]

E' importante sottolineare che i neuroni svolgono le loro attività parallelamente a tutti gli altri e che la conoscenza è distribuita su più neuroni, cosicché anche se uno di essi viene a mancare la rete riesce comunque ad adempiere alle sue funzioni.

1.2

Reti Neuronali Artificiali

Le reti neurali artificiali, **ANN** (“*artificial neural network*”) , sono ispirate a quelle biologiche, ma ovviamente sono molto più semplici e schematiche. I neuroni sono simulati da piccole unità di calcolo (“*units*”) . Ciascuna di esse svolge semplici

computazioni e calcola di volta in volta un nuovo livello di attivazione che propaga attraverso i suoi collegamenti di uscita. Il **calcolo** dipende dai segnali di ingresso e dai pesi (“*weights*”) dei collegamenti di ingresso e si suddivide in due fasi; nella prima si applica la **funzione di ingresso** di tipo lineare che effettua la somma pesata degli ingressi, nella seconda si utilizza una **funzione** detta **di attivazione** di tipo solitamente non lineare che trasforma la somma pesata nel valore di attivazione. La modalità di attivazione della rete può essere sincrona, quindi i neuroni cambiano stato tutti insieme, oppure asincrona, così chiamata perchè i neuroni cambiano stato sequenzialmente secondo criteri stabiliti.

Le architetture principali delle reti neurali sono quella **completamente connessa** e quella **stratificata**.

In una rete stratificata le unità sono disposte a livelli e si classificano in tre gruppi principali: le **unità di tipo input** che ricevono i dati di ingresso e formano lo strato di ingresso, le **unità di tipo output** che forniscono i risultati della rete e formano lo strato di uscita e le **unità di tipo hidden** che, collegate sia agli input che agli output, fungono da intermediari calcolando risultati parziali e formano lo strato nascosto.

Quindi per creare una rete si definiscono:

-il numero dei canali di ingresso: n ;

-il numero dei canali di uscita: m ;

-il tipo di segnali di ingresso: x_j ;

-il tipo di segnali di uscita: y_j ;

-i pesi delle connessioni: w_j ;

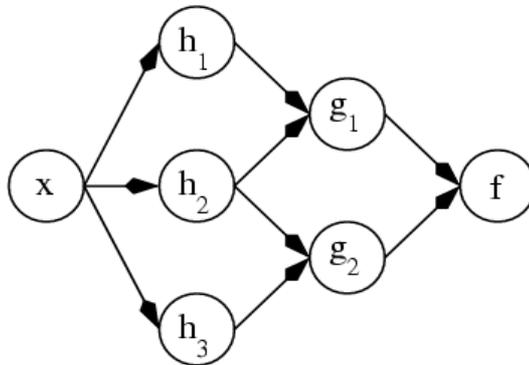
-la funzione di attivazione: F (*alcuni esempi: la funzione gradino, la funzione segno, la funzione sigmoide*);

-la funzione di uscita: f . [001]

Il concetto di rete neurale comporta una decomposizione della funzione f che a sua volta è formata da altre funzioni g che a loro volta possono essere formate da funzioni h e così via. Questo concetto è sintetizzabile con la seguente formula.

$$f(x) = k \left(\sum_j w_j g_j(x) \right)$$

E schematizzabile con la seguente figura.



La figura può essere interpretata sotto due punti di vista entrambi corretti. Il primo, funzionale, secondo cui l'ingresso è trasformato in un vettore a tre dimensioni, che a sua volta si trasforma in un vettore a 2 dimensioni ed infine confluisce nell'uscita.

Il secondo, probabilistico, vede f come una variabile causale che dipende dalle variabili g che dipendono a loro volta dalle variabili h .

Ogni livello della rete in questo modo è indipendente e le computazioni possono avvenire in parallelo. La rete dell'esempio è del tipo **feedforward** che sarà approfondito in seguito. Esistono, però, anche reti che presentano percorsi di connessioni ciclici e sono chiamate per questo **ricorrenti o feedback**. Esse presentano dunque una retroazione per cui la funzione f è dipendente su se stessa.

Al fine di costruire una rete neurale in seguito alla definizione del modello dei neuroni, dell'architettura della rete e della modalità di attivazione dei neuroni vanno scelti il paradigma e la legge di apprendimento.

Data una classe di funzioni che risolvono il problema per cui la rete neurale è stata costruita e data una funzione-costo che quantifica con un numero reale il costo dell'algoritmo, esiste una funzione della classe definita **ottimale**, ossia che risolve il problema nel modo più efficiente possibile minimizzando la **funzione-costo**.

Per i problemi la cui soluzione dipende da alcuni dati il costo è definibile solo in

seguito ad alcune osservazioni, in altri casi invece è possibile definire funzioni-costo ad hoc.

L'apprendimento è definibile come la ricerca di una funzione che si avvicina il più possibile alla funzione ottimale.

Esistono tre grandi paradigmi di apprendimento: quello supervisionato, quello non supervisionato e quello per rinforzo. [010]

L'**apprendimento supervisionato** è una tecnica di apprendimento automatico, in cui vengono definiti un insieme di dati in ingresso I , uno di dati in uscita O e una funzione h che associa ad ogni elemento dell'insieme I un elemento dell'insieme O .

Questi dati, chiamati *training set*, sono utilizzati come esempi per l'algoritmo di apprendimento che, presupponendo che ingressi simili corrispondano ad uscite simili, ricava una funzione h_l che approssima la funzione h . È di fondamentale importanza trovare il **giusto numero di esempi** da fornire alla rete. Pochi potrebbero essere insufficienti per ricavare una funzione adeguata, troppi al contrario potrebbero rallentare l'algoritmo. [011]

L'**apprendimento non supervisionato** modifica i pesi della rete basandosi esclusivamente sui dati in input senza conoscere gli output corrispondenti. Si basa su algoritmi che tentano di classificare i dati in ingresso, raggruppandoli nei cosiddetti *cluster* con criteri topologici o probabilistici.

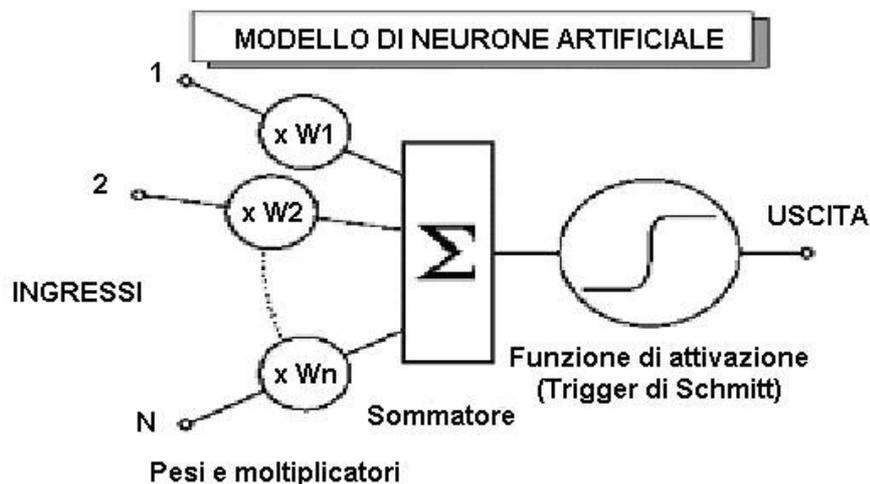
L'**apprendimento per rinforzo** è una tecnica che sfrutta algoritmi in grado di adattarsi e modificare il loro modo di operare in base agli stimoli dell'ambiente esterno al sistema. Una scelta di cambiamento corretta comporterà un premio, mentre una scelta scorretta provocherà prestazioni inferiori. Naturalmente l'obiettivo è massimizzare il premio. Esistono algoritmi di questo tipo ad apprendimento continuo, caratterizzati da una continua retroazione che corregge l'algoritmo ogni istante. Ne sono esempi i programmi del riconoscimento del parlato ed i sistemi OCR (*“Optical Character Recognition”*).

Altri algoritmi di questo tipo invece sono ad addestramento preventivo, ossia dopo aver ricevuto una serie di stimoli dall'esterno e aver trovato una configurazione

ritenuta efficiente, cristallizzano il loro stato rimanendo immutabili. Questa perdita di dinamicità sicuramente comporta come vantaggio un minor uso di risorse e di CPU.[012]

1.3

La Rete Feedforward



Per comprendere questa tipologia di reti si può partire con l'analizzare il modello di neurone artificiale realizzato da **McCulloch e Pitts** nel lavoro del 1943: "*A logical calculus of the ideas immanent in nervous activity*".

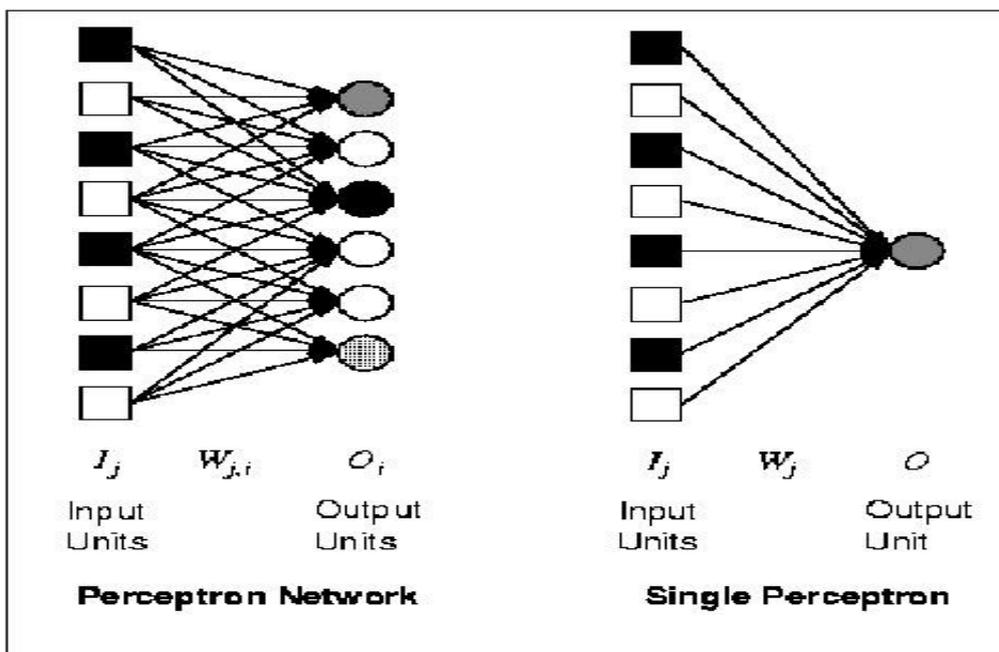
Questo modello è basato sulle seguenti assunzioni fondamentali:

- il neurone è un dispositivo binario, nel senso che può trovarsi in soli due possibili stati: "eccitato" o "non eccitato";
- ogni neurone ha una soglia fissa di eccitazione, oltre la quale esso cambia di stato e manda un impulso elettrico lungo l'assone ;
- lo stato di eccitazione dipende da impulsi ricevuti da sinapsi eccitatorie;

- lo stato di eccitazione dipende anche dagli impulsi ricevuti da sinapsi inibitorie (se una di esse è attiva, il neurone non può attivarsi);
- c'è un tempo finito d'integrazione degli impulsi in ingresso, durante il quale il neurone calcola il suo stato. [013]

I neuroni ricevono degli stimoli e li elaborano. La computazione può essere anche molto complessa, ma è schematizzabile dicendo che gli impulsi in ogni input vengono moltiplicati con il peso della corrispondente connessione e i prodotti sono sommati tra loro. Se la somma risultante supera una certa soglia definita il neurone si attiva. Maggiore è il peso di una connessione più essa risulta importante e costituisce un percorso preferenziale di collegamento ingresso-uscita. Tuttavia bisogna specificare che in misura proporzionale al peso tutte le connessioni contribuiscono al collegamento ingresso-uscita.

Le “reti alimentate in avanti” stratificate sono studiate per la prima volta alla fine degli anni '50 da R. Rosenblatt che introduce lo schema del **perceptron** (**perceptrone**), nome ancora oggi utilizzato per indicare reti feedforward senza strati hidden, ossia le più semplici costruibili.



Una qualsiasi uscita di questo tipo di reti è indipendente dalle altre, poiché ciascun peso influenza solo uno degli output. [014]

Il perceptrone costituisce un progresso decisivo rispetto al modello binario di McCulloch e Pitts, perché i suoi pesi sinaptici sono variabili e quindi esso è in grado di apprendere.

Il modello di apprendimento proposto da Rosenblatt è basato su una funzione di minimizzazione dell'errore che altera i pesi delle connessioni in base alla differenza tra l'uscita effettiva e quella desiderata. [015]

Questa funzione è chiamata **error back-propagation** (retro-propagazione dell'errore) e sarà descritta nel dettaglio nel seguente paragrafo.

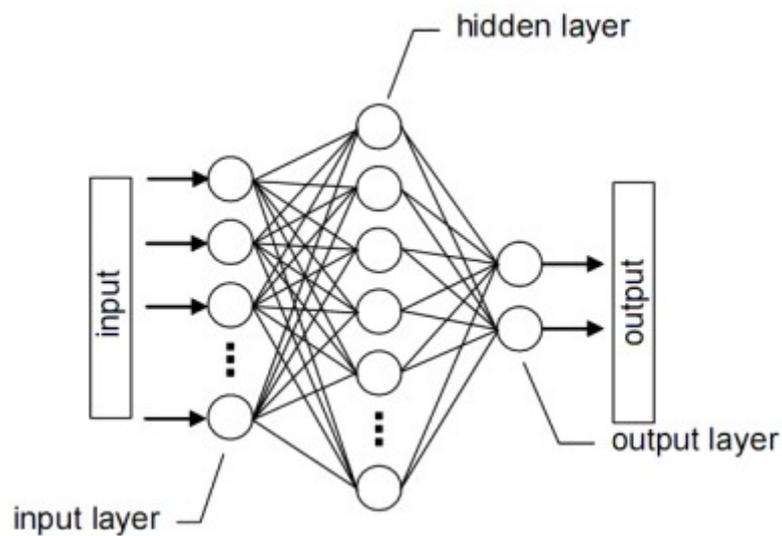
Data la sua semplice struttura definibile con la seguente equazione lineare:

$$O_j = \text{gradino} \left(\sum_j w_j I_j \right) ,$$

il perceptron può rappresentare solo funzioni booleane linearmente separabili, infatti è in grado di realizzare la AND e la OR, ma non la XOR. Questa proprietà è scoperta da Marvin Minsky e Seymour Papert e pubblicata nel 1969 nell'opera "*An introduction to computational geometry*". Tutto ciò ridimensiona l'enorme successo che il perceptron ha riscosso dopo la pubblicazione di Rosenblatt. [016]

Tuttavia negli anni successivi le reti neurali si sviluppano e recuperano gran parte del successo perduto, a partire dal lavoro di Paul Werbos nel 1974 e successivamente di J.J. Hopfield nel 1982 con lo studio dei **MLP** (*Multi-Layers Perceptron*, ossia perceptrone multistrato) .

Tutt'oggi si realizzano reti neurali multistrato feedforward, così chiamate perché non presentano retroazione a differenza delle feedback. Esse sono pienamente connesse, ossia ogni elemento di uno strato è collegato a tutti gli elementi dello strato successivo. I neuroni prendono i loro input solo dallo strato precedente ed inviano i loro output solo allo strato seguente. [017]



Lo strato input si occupa di prendere gli ingressi e se necessario li tratta in modo da adeguarli alle richieste dei neuroni. Lo strato hidden si preoccupa dell'elaborazione vera e propria e può essere composto anche da più colonne di neuroni. Lo strato output raccoglie i risultati e se necessario li adatta e tratta per le computazioni successive.

Gli strati hidden possono essere più di uno e non c'è un limite stabilito insuperabile.

E' stato dimostrato, però, che se gli strati hidden sono più di 1 o al massimo 2 le prestazioni di generalizzazione della rete non migliorano; quindi solitamente dal terzo strato in poi eventuali aggiunte sono ridondanti.

L'efficacia di generalizzazione di una rete neurale multistrato dipende invece in modo significativo dall'addestramento che ha ricevuto e dal fatto di essere riuscita o meno ad entrare in un buon minimo locale. [018]

1.4

L'algoritmo di Backpropagation

Uno dei metodi più noti ed efficaci per l'addestramento delle reti feedforward è il

cosiddetto algoritmo di retropropagazione dell'errore (error backpropagation), proposto nel 1986 da David E. Rumelhart, G. Hinton e R. J. Williams. Tale lavoro fu prodotto riprendendo il modello creato da Werbos. L'algoritmo di *backpropagation* (BP) è una tecnica d'apprendimento tramite esempi, costituente una generalizzazione dell'algoritmo d'apprendimento per il perceptrone sviluppato da Rosenblatt alla fine degli anni '50 che, come detto, era in grado di rappresentare unicamente funzioni booleane linearmente separabili. [019]

L'algoritmo BP insieme alle reti MLP è stato il protagonista del rilancio delle reti neurali dopo il drastico calo di popolarità dovuto alla pubblicazione Minsky e Papert, poiché riesce anche nella rappresentazione di funzioni non lineari quali lo XOR.

Il BP si inserisce nell'apprendimento supervisionato e consente di modificare i pesi delle connessioni al fine di minimizzare una funzione errore definita E .

Per comprendere il funzionamento di questo algoritmo definiamo il vettore h-esimo di output out^h , il vettore h-esimo di ingresso x^h e il vettore h-esimo di output che noi desideriamo y^h il quale insieme ai corrispondenti ingressi fa parte del set di esempi di addestramento chiamato più propriamente **training set**. Il training set è dunque un insieme di N coppie di vettori (y^h, x^h) , con $h = 1, \dots, N$.

La funzione E da minimizzare è scrivibile come:

$$E(w) = (1/2) \sum_h \sum_k (out_k^h - y_k^h)^2$$

dove k rappresenta il k-esimo neurone di output. Per minimizzare $E(w)$, dipendente dai pesi della rete che variano nel tempo, si utilizza il metodo della **discesa del gradiente (gradient-descent)**. Questo algoritmo parte da un punto $x(0)$ e calcola il gradiente $\nabla f(x^0)$. Il gradiente indica la direzione verso cui si massimizza la funzione, per trovare la direzione in cui si minimizza si deve quindi considerare $-\nabla f(x^0)$. Definita la direzione ci si muove di una distanza η definita a priori e si trova quindi un diverso punto $x(1)$. Sul punto $x(1)$ si calcola nuovamente il gradiente, e si procede allo stesso modo iterativamente trovando altri punti e corrispondenti

gradienti fino a quando $\nabla f(\underline{x}^n)=0$.

L'algoritmo BP può essere schematizzato in due fasi. Nella prima detta **Forward Pass** i vettori in input sono applicati ai nodi in ingresso della rete con una propagazione in avanti dei segnali attraverso ciascun livello. In questa fase i pesi delle connessioni sono fissati. Nella seconda fase detta **Backward Pass** la risposta della rete viene confrontata con l'uscita desiderata ottenendo il segnale d'errore che viene poi propagato in verso opposto rispetto alla fase Forward Pass.

Per addestrare una rete neurale con apprendimento supervisionato è necessario:

- creare un training set, ossia un insieme di pattern di input correlati a determinati pattern di output;
- inizializzare i pesi della rete a valori casuali, piccoli rispetto a quelli che assumeranno, ed a norma nulla;
- ripetere iterativamente il ciclo di apprendimento feed forward fino a quando la funzione errore non ha raggiunto un valore soddisfacente o fino a quando non è stato eseguito un numero di iterazioni stabilito a priori.

Il ciclo consiste, innanzitutto, nella scelta a caso di uno dei pattern dal training set. Va specificato che nei cicli successivi esso non potrà essere scelto nuovamente per l'addestramento fino a quando tutti gli altri non saranno stati impiegati almeno una volta. Successivamente si calcolano i valori dei neuroni successivi attraverso sommatorie di produttorie per poi, in ogni neurone, detrarre dal risultato il valore di soglia proprio di quel neurone. Tale differenza sarà poi il segnale propagato al livello successivo, tramite una funzione logistica.

In seguito si confrontano i risultati calcolati con quelli nel training set dei pattern relativi all'input del ciclo. Poi inizia la vera e propria backpropagation dagli output verso gli input in cui si calcola la correzione da apportare ai pesi secondo la localizzazione del minimo scelta e la si applica. [020]

Oltre al training set si costruisce un **validation set** costituito anch'esso da input e relativi output per testare la rete, successivamente all'applicazione dell'algoritmo BP, e verificare il livello di apprendimento.

1.5

Intelligenza Artificiale e Informatica Musicale : Simbolismo e Connessionismo a confronto

Esistono negli ambiti delle scienze cognitive e dell'intelligenza artificiale due paradigmi centrali: il **simbolismo** ed il **connessionismo**. Quest'ultimo impiega le reti neurali sopra descritte.

Il simbolismo basa la computazione sul metodo standard di Von Neumann sotto il controllo di una singola potente unità di calcolo, mentre il connessionismo adotta uno stile che richiama il funzionamento del cervello caratterizzato da una simultanea interazione di un gran numero di piccole unità elementari di calcolo. [021]

I sistemi artificiali simbolici funzionano utilizzando catene di inferenze con variabili logiche, mentre le ANN applicano regole evolutive a variabili numeriche.[022]

Le caratteristiche dei primi sono dunque la logica, la serialità e l'induzione formale. Essi operano ad un alto livello concettuale usando simboli dedicati per rappresentare ogni concetto, adottando un **tipo di rappresentazione** che possiamo definire “**locale**” (*local representation*).

Le caratteristiche delle seconde invece sono la dinamica, il parallelismo e l'utilizzo di esempi. Il loro tipo di rappresentazione è prevalentemente “**distribuito**” (*distributed or subsymbolic representation*).

Premesso che le ANN possono utilizzare entrambi i tipi di rappresentazioni e che essi non sono tra loro esclusivi ma spesso compresenti all'interno di un sistema ecco un esempio della differenza. In ambito di informatica musicale se si usa la rappresentazione locale si designa un neurone per identificare ogni accordo, invece una rappresentazione più distribuita si ottiene modellando ogni accordo mediante le note che lo compongono.

I vantaggi essenziali di un approccio che abbraccia il connessionismo sono i seguenti.

- Il primo è **la capacità di apprendere**. Le ANN solitamente non sono programmate, ma addestrate tramite algoritmi di apprendimento. Questo riduce la complessità dovuta all'ingegnerizzazione della conoscenza ("*knowledge engineering*") del sistema in questione. In particolare per applicazioni musicali il "*knowledge engineering*" è spesso innaturale, difficile e fonte di errori. Per esempio l'opera di composizione musicale sarebbe fortemente ostacolata se costretta in un sistema di regole predefinito e rigido.
- Il secondo è **la capacità di generalizzare**. Grazie ad alcune feature le ANN possono, ad esempio, riconoscere il timbro di uno strumento a prescindere dalla tecnica con cui esso venga suonato, nonostante siano presenti significative differenze nei dati analizzati.
- Il terzo è **la tolleranza ai rumori e alle contraddizioni**. In musica spesso non si seguono regole facilmente formalizzabili e questa caratteristica è un forte vantaggio ai fini della modellazione dei sistemi.
- Il quarto è **la tolleranza verso un sovraccarico di informazioni**. Le ANN non hanno una memoria fissa e quando si presenta un eccesso di informazioni le reti neurali tendono a raggruppare i dati simili, applicando una generalizzazione delle feature.

Il connessionismo d'altra parte comporta anche diversi problemi e svantaggi, riassunti di seguito.

- Il primo è **la limitazione a piccoli problemi**. Per lo meno nell'ambito musicale l'impiego di ANN fino ad oggi è stato limitato a problemi di piccole dimensioni e stilizzati. La ricerca in questo campo è ancora in una fase iniziale.
- Il secondo è **la difficoltà a realizzare lunghe catene di inferenze**. In questi casi l'approccio del simbolismo è indiscutibilmente più efficace.
- Il terzo è **la limitata capacità di spiegare il processo attraverso cui si arriva ai risultati**. Essendo le ANN basate in gran parte su algoritmi stocastici ed imitativi di esempi è molto difficile spiegare come si arrivi a certi livelli di apprendimento. Una

ANN può essere addestrata ad armonizzare melodie in un dato stile, ma è difficile analizzare in che modo la rete arrivi ad una particolare soluzione.

-Il quarto è **la difficoltà a rappresentare la conoscenza in modo strutturato e a modellarla mediante metalivelli. [023]**

2.0

Percezione Sonora e Musica

Per costruire una macchina che imita attitudini umane può essere di grande aiuto lo studio dei meccanismi di percezione dell'uomo. In questo capitolo inizialmente si dà una definizione generale di percezione per poi scendere più nel particolare descrivendo quella sonora ed i principi di base della psicoacustica. In seguito si definiscono il concetto di nota e di timbro, discutendo la sua rappresentazione statica e la sua fenomenologia temporale. Si sale poi di livello analizzando la musica in tre dei suoi aspetti fondamentali l'armonia, il ritmo e la melodia. Segue un paragrafo che espone qualche illusione ed effetto sonoro e musicale. Esso è una riflessione sul fatto che gran parte di ciò che udiamo, oltre ad essere profondamente influenzato dalla conformazione dell'apparato uditivo, è il frutto di una rielaborazione del cervello condizionato da pregiudizi ed aspettative. Infine alla luce di quelli che sono a grandi linee i meccanismi di percezione, la struttura dei suoni e gli aspetti principali di una composizione musicale si analizzano le differenze nel riconoscimento di pattern musicali tra uomo e macchina, importanti per affrontare il tema di classificazione di file audio tramite reti neurali presente nel capitolo seguente.

2.1

Percezione uditiva umana

L'uomo è un elaboratore di informazioni. La scienza che si occupa dei processi attraverso i quali egli recepisce, analizza, immagazzina ed assimila i dati è la

psicologia cognitiva. L'informazione viene processata attraverso diversi stadi. Inizialmente sono coinvolti i registri sensoriali che, in base al livello di attenzione dell'individuo, rilevano ed immagazzinano una serie di pattern. Questi sono confrontati con pattern presenti in archivi della memoria a breve o lungo termine e sono aggiunti a loro volta a tali archivi. Il confronto e l'analisi dei pattern fanno scaturire una eventuale decisione da parte dell'individuo che consiste in una selezione di una risposta alle informazioni ricevute. L'elaborazione delle informazioni è costituita da due aspetti inscindibili; quello che concerne i dati rilevati dall'esterno e quello che riguarda i concetti mentali interni già presenti, i quali da un lato aiutano a percepire la realtà perchè la organizzano secondo criteri rappresentativi, dall'altro la modificano modellandola secondo la loro struttura.

La **percezione** è definibile come l'organizzazione immediata e dinamica delle informazioni sensoriali, scaturite da un insieme di stimoli, delimitata nello spazio e nel tempo. Dalla percezione scaturiscono le **sensazioni**, ossia immediate impressioni soggettive che corrispondono ad una determinata intensità dello stimolo fisico.

E' importante sottolineare che percepiamo solo le forme di energia che i nostri ricettori riescono a rilevare. I **ricettori** sono cellule specializzate che traducono l'energia giunta agli organi di senso in segnali elettrici che raggiungono il cervello.

La disciplina che si occupa delle variazioni delle sensazioni in base alla variazione degli stimoli appena descritti è la **psicofisica**. [024]

In particolare la **psicoacustica** è la branca della psicofisica che studia i suoni.

Nonostante le sensazioni provocate dai suoni siano differenti da individuo ad individuo si trovano anche molte ricorrenze nelle reazioni a determinati stimoli sonori che costituiscono un interessante oggetto di ricerca.

Il suono è prodotto dalle vibrazioni di un mezzo elastico. Le onde sonore possono essere descritte in termini di **frequenza** ed **ampiezza**. Un suono puro, non udibile in natura, ma riproducibile artificialmente è un'onda sinusoidale di frequenza costante. Si misura in hertz (Hz), dove 1 Hz caratterizza un "evento" che occorre una volta in un secondo. I suoni reali sono una sovrapposizione di un numero infinito di onde con

frequenza ed ampiezza variabili. L'ampiezza definisce la variazione massima della pressione dell'aria. La pressione si misura in Pascal ma più spesso si fa riferimento all'ampiezza del suono in rapporto alla soglia uditiva dell'orecchio, il rapporto si esprime in decibel (dB).

Il suono dunque è un segnale analogico continuo che può portare un numero teoricamente infinito di informazioni.

L'orecchio umano può udire dai 20 Hz ai 20 KHz. Con l'avanzare dell'età questo intervallo si restringe, soprattutto il limite superiore che può scendere fino a 16 kHz negli adulti. Alcune frequenze al di fuori dell'intervallo indicato possono essere percepite attraverso il tatto.

Per misurare l'effetto della frequenza sull'orecchio si segue una scala logaritmica. L'altezza percepita di un suono è funzione esponenziale della frequenza. La scala musicale più usata in occidente, quella del temperamento equabile è un esempio chiarificatore di questa relazione. Quando la frequenza fondamentale di una nota è moltiplicata per $2^{1/12}$ il risultato è la frequenza della nota un semitono sopra.

Considerando l'intensità del suono l'intervallo va da un minimo di 0 dB a un massimo che possiamo identificare con la soglia del dolore. Il dolore oltre che dall'intensità dipende anche dalla durata, un suono istantaneo a 120 dB può essere innocuo, ma uno prolungato a 80 dB può provocare sordità permanente.

In realtà la soglia minima di udibilità dipende dalle frequenze. L'orecchio denota il picco di sensibilità in un intervallo che si aggira tra 1 kHz e 5 kHz, ma le soglie possono cambiare con l'età.

La presenza di due orecchie in posizioni opposte del cranio, caratterizzate da una differenza di fase nel processo uditivo della stessa fonte sonora, è fondamentale per elaborare la **localizzazione del suono**.

Quando un suono più debole è reso inudibile da un altro più forte si parla di **mascheramento**. In caso in cui i suoni siano simultanei questo fenomeno avviene in modo più pronunciato per suoni della stessa frequenza e si parla di **mascheramento di frequenza**. Se invece un suono debole segue immediatamente o anche precede di

poco un suono molto forte può verificarsi il **mascheramento temporale**. [025]

2.2

Note e Timbro

Gli strumenti musicali generano dei toni che li contraddistinguono. L'istanza di un tono eseguito per una certa durata, mantenendo costante la sua altezza, prende il nome di **nota**. Essendo costituiti da corpi vibranti gli strumenti possono produrre diverse frequenze. Ogni tono è costituito da un numero di frequenze distinte, la più bassa delle quali è chiamata **fondamentale**, mentre le altre prendono il nome di **armoniche**. Le armoniche di intensità significativa sono essenziali per la caratterizzazione del timbro. [026]

Una definizione diffusamente accettata di timbro è la seguente: il **timbro** è la qualità percepita di un suono che ci permette di distinguere due suoni che hanno la stessa altezza e la stessa intensità.

Ragionando da un **punto di vista statico** si può dire che il timbro di uno strumento è dovuto, in larga parte, alla composizione spettrale del suono che esso emette. A causa della particolare geometria delle componenti vibranti specifiche, la configurazione della fondamentale e delle armoniche generate varia da strumento a strumento.

La somma delle frequenze armoniche e della fondamentale da origine a **forme d'onda** diverse e caratterizzanti.

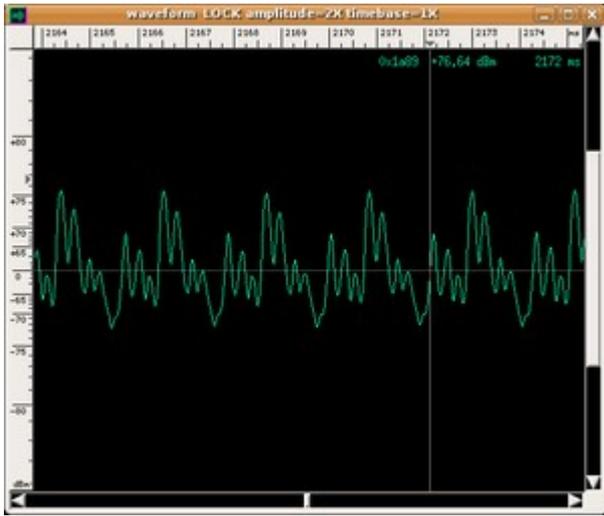


Figura: Forma d'onda del Violino

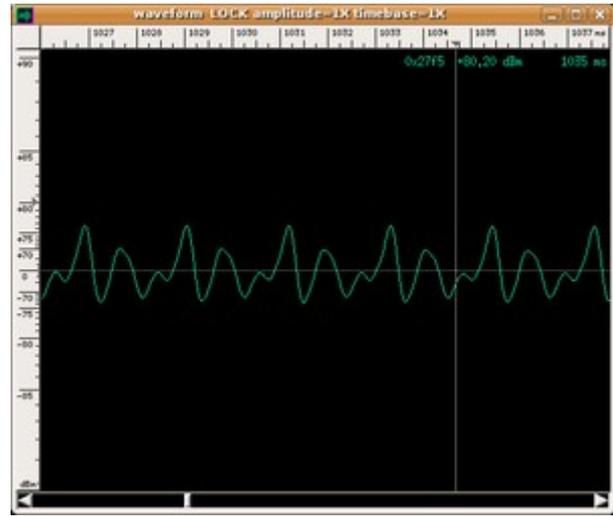


Figura: Forma d'onda del Flauto

Gli spettri che rappresentano queste forme d'onda sono, però, insufficienti per spiegare la complessa percezione del timbro. Il suono reale è qualcosa di vivo e dinamico ed è necessario analizzarlo da un **punto di vista dinamico**. Esso infatti segue una parabola costituita da fasi temporali ben precise che testimoniano una evoluzione dell'ampiezza dell'onda sonora nel tempo.

La prima è la fase di **attacco** (*attack*) che dura fino a quando il suono ha raggiunto l'ampiezza massima. La sua durata varia da strumento a strumento. Per fare alcuni esempi nelle percussioni è molto breve, negli strumenti ad arco invece varia a discrezione dell'esecutore.

La seconda fase, chiamata **decadimento** (*decay*). E' presente in quegli strumenti in cui per produrre il suono desiderato è necessario superare una soglia fisica oltre la quale esso scatta. Un esempio eclatante è la tromba in cui l'esecutore dopo aver superato la soglia di pressione d'aria necessaria corregge la forza del soffio al fine di raggiungere la fase di regime.

Il terzo stadio, di regime appunto, è chiamato **tenuta** (*sustain*). Il suono rimane stabile mentre l'esecutore continua a fornire energia; corrisponde alla rappresentazione delle forme d'onda descritte nell'analisi statica. Negli strumenti a percussione la tenuta praticamente non esiste.

La quarta ed ultima fase è il **rilascio**; inizia quando l'esecutore smette di fornire

energia e l'ampiezza decade più o meno rapidamente a seconda degli strumenti. Nelle percussioni ad esempio solitamente è lunga, così come nel pianoforte o nella chitarra, mentre è più corta negli strumenti ad arco o a fiato. [027]

2.3

Armonia, Ritmo e Melodia

Armonia, ritmo e melodia sono i tre pilastri su cui si basa ogni composizione musicale.

Nel paragrafo precedente sono state definite le note in senso isolato come toni di una determinata frequenza. Per descrivere i tre pilastri sopracitati è necessario parlare di relazioni tra note.

L'**armonia** è il ramo della teoria musicale che studia la sovrapposizione “verticale”, ossia simultanea, dei suoni. [028]

Il **ritmo** è il susseguirsi di una serie di accenti con una periodica regolarità. Esso è basato sulla suddivisione del tempo in forme e misure variabili, talvolta regolari e simmetriche altre volte irregolari e asimmetriche. Il ritmo è quindi una struttura dinamica che si ripete ciclicamente, e le diverse tipologie di struttura sono chiamate **metri**. [029]

La durata delle insorgenze di note che vengono suonate in successione è quello che produce un senso del ritmo. [026]

La **melodia** è una successione di note di diversa altezza e durata che si sviluppano nel tempo, la cui struttura forma un organismo musicale di senso compiuto. [030]

La fondamentale relazione tra due note distinte è l'**intervallo**, e definisce la loro distanza. A causa della fisiologia della percezione uditiva l'intervallo non è proporzionale alla differenza tra le frequenze dei due suoni, ma al loro logaritmo quindi al loro rapporto. Ci sono due tipologie di intervallo quello **armonico** e quello

melodico. L'armonico può essere caratterizzato da consonanza o dissonanza in base a quali sono le armoniche comuni tra i due suoni. Il melodico è classificato come discendente o ascendente in relazione all'altezza della seconda nota rispetto alla prima in senso temporale. [031]

Queste componenti musicali sopradescritte hanno subito nel corso della storia grandi mutamenti ed evoluzioni. La ripetizione e la simmetria di gruppi di note caratterizzati da aspetti armonici, ritmici o melodici comuni costituisce **pattern musicali** che contraddistinguono i compositori appartenenti a differenti culture e periodi storici.

Per esempio la musica Occidentale contemporanea utilizza una **scala** di 12 note e tende verso un' alta complessità sia armonica, sia ritmica che melodica, ma non è sempre stato così.

Nel primo medioevo le composizioni sono per lo più monofoniche con accompagnamenti secondari. La polifonia emerge nel Rinascimento e culmina nel periodo Barocco con la tecnica del **contrappunto**. In esso coesistono diverse voci indipendenti che si intersecano creando interazioni tra le linee melodiche.

Verso la fine del Barocco il contrappunto perde gradualmente la sua posizione predominante, lasciando sempre più il posto a composizioni governate da principi armonici. E' così che cominciano a diffondersi gli **accordi**, ossia insiemi di 3 o più note suonate simultaneamente. I pattern delle progressioni di accordi sono fondamentali nella musica Occidentale, dal periodo Barocco al Jazz moderno. [026]

Gli accordi e i **modi musicali**, ossia insiemi di intervalli derivati da una corrispondente scala musicale variando semplicemente la nota iniziale, sono inoltre di fondamentale interesse perchè costituiscono i principali strumenti attraverso cui il compositore comunica le sue impressioni all'ascoltatore. [032] Infatti facendo un semplice esempio pratico è noto che la predominanza di accordi minori trasmette sensazioni di malinconia ed in alcuni casi tristezza, mentre la prevalenza di accordi maggiori può suscitare felicità od addirittura euforia. Ovviamente le reazioni dipendono in certa misura dalla soggettività dell'ascoltatore. La quantità di impressioni suscetibili con diverse progressioni di accordi è vasta forse quanto la

gamma delle emozioni umane.

2.3

Illusioni sonore

Il sistema sensoriale di regola è molto efficiente ma può essere “ingannato”.

La conoscenza preacquisita e le aspettative influenzano la percezione significativamente.

Citando il neuroscienziato David Marr: “Percepire è costruire un significato” (David Marr, 19 Gennaio, 1945 - 17 Novembre, 1980). [024]



Figura: “L'uomo con il cubo” M.C. Escher

Nella percezione uditiva interagiscono in modo complesso e delicato le caratteristiche

fisiche del suono ricevuto, quelle fisiologiche dell'orecchio e l'attività neurale del cervello. E' errato considerare ciò che si ode come un'immagine fedele del suono percepito, essa è piuttosto una sua elaborazione spesso non fedele all'originale.

Ne sono una chiara dimostrazione gli **effetti**, definibili come percezioni "non fedeli" dovute alle caratteristiche fisiologiche dell'apparato ricevente, e le **illusioni**, ossia il prodotto dell'elaborazione del cervello che interpreta i suoni che riceve.

Una volta che si è presa consapevolezza delle illusioni, i loro meccanismi possono, però, rivelarsi un utile strumento di studio.

La **scala di Shepard** è una delle illusioni più diffuse. Se ascoltiamo un campione periodico in cui le armoniche più acute vengono reintrodotte gradualmente, mentre quelle più gravi svaniscono in modo costante, si sente un tono discendente senza fine. In realtà, come detto, il tono è sempre lo stesso, ma cambiano le armoniche.

Questo è un esempio di effetto di tipo Shepard, esso è molto usato anche in musica. Per citare solo alcuni gruppi famosi, è stato usato dai Queen e dai Pink Floyd. Persino Bach aveva studiato questo fenomeno utilizzandolo sotto forma di progressioni modulanti nella *Fantasia e Fuga in Sol minore* per organo BWV 542.

Anche la modalità in cui percepiamo **la fondamentale** di una nota riserva delle sorprese. Infatti se si ascoltano due successioni melodiche identiche in ogni aspetto, ma in una delle due si tolgono tutte le fondamentali lasciando solo le armoniche, sorprendentemente le note delle due melodie paiono della stessa altezza anche se sembra cambiare il timbro dello strumento esecutore. Questo esperimento ci suggerisce che il cervello attribuisce l'altezza di un suono basandosi più sui rapporti tra le armoniche di tutto lo spettro, che non sulla sola fondamentale.

E' possibile udire suoni "fantasma". Questo è ciò che avviene quando si ode il cosiddetto **terzo suono di Tartini**. Il suono "fantasma" si percepisce quando due suoni intensi e ricchi di armonici giungono simultaneamente all'orecchio. E' piuttosto comune sentire questo effetto con il violino suonando sulle prime due corde note doppie. Non è un caso, infatti, che lo scopritore del terzo suono, Giuseppe Tartini, fosse un eccelso violinista di fine '700. Il fenomeno è stato interpretato come una

prova del comportamento non lineare dell'orecchio umano, che per intensità elevate del segnale in ingresso lo può distorcere. Si creano quindi realmente nell'orecchio interno massimi fisici dell'onda di pressione cocleare. La distorsione in questo caso è quella di intermodulazione che produce armoniche pari alla somma e alla differenza tra le frequenze dei suoni in ingresso.

Un altro interessante effetto è l'**effetto McGurck** che riguarda la complementarità tra la percezione visiva e quella uditiva. Osservando il labiale del nostro interlocutore è frequentissimo farci influenzare dal movimento delle sue labbra per interpretare quello che sta dicendo. Se un video mostra un uomo che dice “ga”, ma in realtà l'audio riproduce “ba”, chi osserva lo interpreta o come una sillaba intermedia come “da” o addirittura la sillaba che suggerisce il labiale “ga”. L'effetto opposto è lo scarso disagio che ci provocano gli asimmetrici movimenti delle labbra nei film doppiati. Questi fenomeni dimostrano che tendenzialmente ricaviamo le informazioni del contesto più facili e comprensibili, tralasciando i particolari incomprensibili che risultano contraddittori con gli aspetti prevalenti. In altre parole, siamo portati a sentire quello che ci aspettiamo di sentire.

E' inoltre significativo analizzare la **differenza tra senso sonoro e senso musicale**.

Il modo in cui il sistema uditivo opera per riconoscere lo strumento (senso sonoro) e il brano in questione (senso musicale) è dimostrabile, grazie a semplici esperimenti, essere diverso. Le informazioni circa lo strumento riguardano la percezione del timbro e si riferiscono alle caratteristiche acustiche dei suoni che compongono il brano. Le informazioni sul brano da un punto di vista musicale, invece, sono contenute nella sua struttura di alto livello, come la sequenza delle note, la melodia, l'armonia, la struttura della composizione: tutte caratteristiche che richiedono una analisi “linguistica” degli stimoli sonori. [[033](#)]

2.4

Differenze nel Riconoscimento di Pattern Musicali tra Uomo e Macchina

L'uomo ha la straordinaria capacità di riconoscere pattern musicali durante l'ascolto di composizioni mai udite e ricondurle allo stile di altri brani che già conosce.

Per un sistema artificiale questo compito è di gran lunga più complesso da realizzare.

La causa principale della differenza tra uomo e macchina nella capacità di riconoscimento è il fatto che i dati in ingresso ad un computer non sono nella loro forma originale, ma necessariamente devono essere rappresentazioni delle informazioni in forma numerica. [026]

Il **riconoscimento di pattern** implica due fasi fondamentali. La prima consiste nell'estrapolare dall'oggetto di indagine delle proprietà significative, chiamate **feature**. La seconda consiste in una **ricostruzione** dell'oggetto analizzato sulla base delle proprietà sopracitate. Il pattern verrà riconosciuto se esiste una elevata somiglianza tra questa ricostruzione ed altre compiute nel passato già classificate secondo certe etichette.

Entrambe le fasi per un sistema artificiale comportano grandi difficoltà.

Per quanto riguarda la prima: quali sono le feature veramente significative? A quale livello di astrazione devono appartenere?

Nel caso della percezione umana questo problema è superato con una **estrapolazione** delle proprietà di tipo **trasversale**, fortemente legata alla soggettività e all'esperienza personale dell'ascoltatore che acquiscono la memoria grazie ad associazioni visive o di altra natura legate ai ricordi. Dunque si mescolano e completano a vicenda sia feature riconducibili al senso sonoro sia al senso musicale, sia ad entrambi. Il cervello compie simultaneamente un'analisi armonica, ritmica e melodica del brano musicale, secondo regole e meccanismi altamente dinamici e difficilmente standardizzabili.

La macchina invece può operare estrazione di feature considerando un livello di

astrazione per volta. Può valutare simultaneamente feature di diverso tipo, riguardanti ad esempio il timbro e la progressione di accordi, ma fatica a rappresentare proprietà frutto della fusione di caratteristiche di basso ed alto livello.

L'implementazione della seconda fase è forse ancora più complicata. La ricostruzione di oggetti operata dal cervello umano è anch'essa dettata da una forte soggettività, ed ogni individuo raggiunge il suo grado di efficienza nel riconoscimento seguendo iter personali. Il punto fondamentale è che il nuovo oggetto mentale costruito non è confrontato con quelli presenti nella memoria seguendo criteri di uguaglianza, ma **criteri di somiglianza**. La ricerca di “oggetti simili” in database virtuali deve fare ancora molti passi in avanti, soprattutto perchè è difficile scegliere quali siano gli **aspetti** su cui misurare il grado di somiglianza e spesso risulta complicato implementare la rappresentazione di questi aspetti.

Un'altra importante e curiosa differenza nel riconoscimento di pattern musicali tra uomo e macchina riguarda le illusioni descritte nel precedente paragrafo. Come si è visto gli effetti e le illusioni svolgono un ruolo centrale nella percezione uditiva e quindi nelle reazioni provocate all'ascoltatore umano. I **computer** in un certo senso sono molto più “**difficilmente ingannabili**” e questo comporta significative conseguenze. Prendiamo come esempio l'effetto Tartini. Se il suono fantasma non esiste, ma è l'ascoltatore che in qualche modo lo “crea”, una macchina non lo potrà mai rilevare. Allo stesso modo, si è descritto il fatto che un suono privato della fondamentale, mantenendo tutte le armoniche, suggerisce all'uomo la stessa altezza con un timbro leggermente diverso. Per una macchina non addestrata a gestire questa illusione la mancanza della fondamentale può essere rilevata come una differenza significativa nel determinare l'altezza del suono.

3.0

Classificazione di File Audio Mediante Reti Neuronali

Questo capitolo inizialmente descrive la classificazione di file audio generici per poi scendere nell'ambito che interessa questa tesi più da vicino, ossia il MIR (*Music Information Retrieval*), il quale si occupa della classificazione e della ricerca di file musicali. Si spiegano i vantaggi di un approccio di analisi basato sul contenuto e non semplicemente sulla nomenclatura dei file. Inoltre si passano in rassegna i campi disciplinari che interessano questo tipo di ricerche; quelli più pragmatici e commerciali, ossia l'efficienza e la qualità di ricerca nei database, ma anche quelli più speculativi ed accademici, cioè la psicologia cognitiva e l'intelligenza artificiale.

Si discutono poi i criteri secondo cui si possono organizzare i database musicali; in particolare le classificazioni per genere e per autore.

Infine si passano in rassegna i metodi di classificazione più diffusi che impiegano reti neurali, descrivendoli in base ai due elementi fondamentali che li contraddistinguono; i formati dei file musicali utilizzati e le diverse tipologie di feature estrapolate.

3.1

Classificazione Audio ed Intelligenza Artificiale

L'uomo ha grande abilità nel riconoscere **il tipo di informazione sonora** che percepisce (voce, rumore o musica), **la velocità** di trasmissione e **il “mood”** di una

musica, ossia se è felice, triste, agitata o rilassata, così come è in grado di giudicare se un brano è simile o dissimile ad un altro.

Come si è visto nel capitolo precedente, invece, per le macchine non è facile implementare metodi di *“query by example”*, un motivo in più rispetto a quelli descritti è che un calcolatore ha a disposizione solamente una serie di *“sample”* (campioni) e un loro confronto diretto è spesso difficilmente realizzabile dato che può variare la *sampling rate* o il *bit rate*. [034]

Tuttavia la classificazione di file multimediali è sicuramente un campo di grande interesse in questa era dell'informazione digitale. I database multimediali dei grandi server, ma anche quelli dei clienti privati che accedono ad Internet, hanno dimensioni gigantesche e sono destinati a crescere. Ecco allora che nasce la necessità di organizzare strutturalmente tali database in modo automatico, senza che il gestore del server o il client in questione siano costretti ad aprire ogni file per metterlo a mano in questa o quella directory. Il vantaggio pratico più immediato di una buona organizzazione dei dati è sicuramente **un aumento della qualità e dell'efficienza della ricerca**. Un database musicale ben strutturato permette di cercare file secondo criteri scelti dall'utente e quindi garantisce probabilità di successo più elevate.

E' nata a questo riguardo una vera e propria sottodisciplina nell'ambito delle scienze dell'informazione multimediale che risponde all'acronimo **MIR** (*Music Information Retrieval*). [035]

Sono state escogitate varie **tecniche di classificazione basate sul contenuto audio** e non meramente sul nome dei file. Una delle tecniche più diffuse ed efficaci consiste appunto nell'utilizzo di reti neurali. I motivi per basare la classificazione sul contenuto sono molteplici; primi fra tutti la maggiore sicurezza ed affidabilità, ma anche la versatilità e l'efficienza. Un esempio banale è che se una canzone è stata chiamata con un nome abbreviato o esso è in una lingua incomprensibile per il database che usa solo tecniche testuali, la canzone non può essere classificata. Un altro esempio significativo riguarda l'ambito del copyright dei file multimediali in rete. Per le case discografiche la classificazione in base al contenuto è utilissima per

assicurarsi che musica protetta da copyright non sia distribuita liberamente in rete. Filtri basati sui nomi dei file si sono rivelati altamente inefficaci; agli utenti basta semplicemente modificare il testo per aggirare tali filtri. Anche modificare il contenuto audio è possibile, ma senza dubbio risulta più complesso e dispendioso.

[036]

L'utilizzo di sistemi intelligenti per l'analisi di file audio non ha solo finalità direttamente legate al miglioramento di filtri o database. L'applicazione di tali tecnologie stimola la ricerca sulla modellazione della percezione uditiva umana.

E' auspicabile infatti che **la psicologia cognitiva e l'intelligenza artificiale** siano discipline che collaborino dialetticamente. Da un lato, infatti, l'intelligenza artificiale si serve delle scoperte biologiche per imitare gli efficienti meccanismi di apprendimento e percezione degli uomini, dall'altro la biologia e la psicologia cognitiva osservano criticamente i risultati ottenuti dai computer per confrontarli con i comportamenti dei viventi e studiare questi ultimi sotto una nuova luce.

Questa cooperazione mutuamente proficua tra le due discipline può aiutare a comprendere quali capacità umane di individuazione di pattern musicali siano realmente automatizzabili e in che misura.

3.2

Criteri di Classificazione

Per implementare tecniche di classificazione o ricerca in base al contenuto è necessario considerare le diverse tipologie di audio quali il parlato, il rumore e la musica. Questa distinzione è di grande utilità perchè le elaborazioni, le indicizzazioni e gli algoritmi applicati variano a seconda delle tipologie e richiedono una gestione separata. Ad esempio per il parlato si usano tecniche di riconoscimento (*speech recognition*) e di indicizzazione delle parole trovate. L'ambito di riconoscimento del

parlato è quello in assoluto più evoluto nel campo dell'informazione audio. Sono state fatte al riguardo numerose ricerche utili allo sviluppo della telefonia, di programmi di dettatura o di autenticazione. Il problema di riconoscimento è un problema di *pattern matching*, che viene affrontato attraverso un **processo statistico** fornendo al sistema una conoscenza del linguaggio utilizzato grazie ad un dizionario dei termini e delle regole grammaticali che determinano le sequenze di parole ammesse. [034]

Anche per quanto riguarda i file musicali si ha a che fare con un linguaggio. Esso possiede una serie di vocaboli, che possono essere identificati con le note ed i timbri, e delle regole grammaticali, quali i metri ritmici e le regole armonico-melodiche. Il linguaggio musicale, però, seppur per molti aspetti sia analogo ad una lingua parlata, risulta essere in generale molto più vario, indefinito e caratterizzato da svariate eccezioni imprevedibili, talvolta anche all'interno di uno stesso ambito musicale.

La complessità della tipologia musicale dei file audio dà adito a molti criteri di classificazione. I più diffusi sono la **classificazione per genere e per autore**.

Le classificazioni per genere aiutano sì a gestire grandi quantità di brani, ma spesso “fanno violenza” alla natura dei brani stessi che, incasellati in certi schemi o diciture, rischiano, se consultati con pregiudizio, di perdere nella fase di ascolto parte della loro complessità e bellezza. Tuttavia sarebbe anche assurdo negare in assoluto l'esistenza dei generi musicali che sono frutto di determinati contesti storico-culturali e facilitano la ricerca e quindi la fruizione delle opere.

La classificazione per autore dal canto suo è sicuramente più fine e richiede una analisi ed un apprendimento di livello più “basso” (particolare) del contenuto dei file da parte dei sistemi artificiali. Questo criterio, però, non va inteso semplicemente come una “classificazione per genere più dettagliata”. Un motivo fra tutti di ciò è che uno stesso compositore potrebbe scrivere brani ascrivibili a generi musicali diversi, mantenendo però alcuni tratti caratteristici che lo distinguono e rendono unico, quindi riconoscibile.

La scelta del criterio di classificazione va fatta in base al contesto, al livello di dettaglio richiesto e alle esigenze del particolare sistema in questione. Inoltre è

auspicabile che i due criteri vengano usati in modo complementare per raggiungere un livello di analisi più completo.

3.3

Metodi di Classificazione:

Formati Audio ed Estrazione di Feature

I metodi di classificazione dipendono da due aspetti fondamentali: il **formato** dei file musicali da analizzare e le **feature** estrapolate dai file stessi.

Le feature, come accennato nel precedente capitolo, sono alcune caratteristiche del file espresse tramite valori numerici.

Il procedimento di gran parte di questi metodi è analogo ed è schematizzabile in tre fasi come segue.

La prima fase consiste nella **estrazione delle feature** dal file. Si ottiene a questo punto un vettore di valori numerici corrispondenti alle diverse proprietà.

Nella seconda fase avviene un **processamento dei valori** ottenuti, che spesso è una normalizzazione o comunque un eventuale aggiustamento dei dati numerici per ricondurli ad una forma decodificabile dalla rete neurale.

Nella terza fase si impiega l'array dei valori processati come **gruppo di pattern di ingresso alla rete**. L'apprendimento, come visto nel capitolo sulle reti neurali, dipende dal numero di esempi del training set, dal tipo di rete e dall'algoritmo di apprendimento impiegato. [036]

I file musicali digitali possono sommariamente essere divisi in due categorie: quelli in un formato che codifica le forme d'onda ("**wave format**") e quelli in un formato che consiste di rappresentazioni di note ("**symbolic format**").

I primi a loro volta si suddividono in file che sono dirette registrazioni di forme d'onda (e.g, **WAV**) ed i restanti che sono il risultato di tecniche di compressione (e.g,

MP3). Nel caso del WAV le registrazioni avvengono tramite microfoni o device simili, sono trasformate in segnali digitali e memorizzate. Gli MP3 invece codificano le informazioni audio cercando di replicare il meno possibile quelle ridondanti e togliendo quelle che sarebbero inudibili per motivi fisiologici o comunque mascherate dagli effetti psicoacustici descritti nel capitolo precedente. Per questi file sono applicati metodi di estrazione che riguardano la **teoria dell'elaborazione dei segnali**. Le bande nel dominio delle frequenze sono computate utilizzando filtri passa-banda che suddividono il segnale in frequenze separate processabili indipendentemente. Questo processo è assimilabile alla separazione delle frequenze che compie il sistema uditivo ed è indicato per ricavare informazioni circa il timbro o il tempo.

I file in formato simbolico contengono informazioni riguardanti la tonalità, l'attacco, la durata e il volume delle note del brano. Questa modalità di rappresentare i dati consente di estrarre informazioni di alto livello sulle proprietà musicali, piuttosto che su quelle sonore. [036]

Alla luce di questa panoramica sui formati musicali si può procedere con una classificazione delle feature. Le **categorie** in cui può ricadere una feature musicale sono essenzialmente le tre descritte di seguito. [037]

- Le **feature di basso livello** (“*low level features*”) nella maggioranza dei casi non procurano informazioni che sembrano intuitivamente musicali, ma hanno una potente capacità interpretativa del brano perchè analizzano direttamente la conformazione del segnale. Le ampiezze dei segnali audio sono rappresentabili nel dominio temporale o nel dominio in frequenza.

Nel **dominio temporale** lo 0 corrisponde al silenzio. L'ampiezza dipende dalla pressione sonora che può essere superiore o inferiore all'equilibrio atmosferico del silenzio. Ogni *sample* è rappresentato da un valore numerico quantizzato. Tra le feature derivanti dall'analisi del dominio del tempo spiccano le seguenti. L' **average energy** è un valore numerico del volume medio del segnale. La formula generale utilizzata per calcolarlo indicando con E l'energia, con N il

numero dei sample e con $x(n)$ il valore del sample ennesimo è:

$$E = (\sum_{n=0}^{N-1} x(n)^2) / N.$$

La **zero crossing rate** indica con quale frequenza il segnale cambia di segno. Essendo $\text{sgnx}(n)$ la funzione segno: $ZCR = (\sum_{n=1}^N |\text{sgnx}(n) - \text{sgnx}(n-1)|) / 2N$. La **silence ratio** è la percentuale del segnale prossima al silenzio. Il silenzio è definibile come un periodo di tempo in cui l'ampiezza è inferiore ad una certa soglia predefinita. Esiste anche una soglia minima temporale per poter definire un certo intervallo come silenzio. Questa feature è il rapporto tra la somma dei periodi di silenzio e la lunghezza complessiva del brano audio.

La rappresentazione del segnale nel **dominio delle frequenze** è la distribuzione dell'audio nelle varie frequenze. E' ottenibile mediante la Trasformata di Fourier e spesso viene chiamata **spettro sonoro**. Più propriamente per le rappresentazioni al computer la trasformata riguarda campioni discreti, quindi prende il nome di **DFT** (*Discrete Fourier Transform*) per il passaggio dal dominio del tempo a quello delle frequenze e di **IDFT** (*Inverse Discrete Fourier Transform*) per il passaggio inverso. Le feature che si possono ricavare dall'analisi dello spettro sonoro sono innumerevoli, di seguito se ne presentano solo alcune.

La **bandwidth** indica il *range* di frequenze occupato da un suono e si calcola come la differenza tra la frequenza massima e quella minima diversa da zero. La **energy distribution** rappresenta la distribuzione del segnale nelle alte o basse frequenze, dove la definizione di alte e basse dipende dall'applicazione. Un importante elemento che funge da indicatore è il **centroide spettrale** che è il punto medio della distribuzione di energia.

La **harmonicity** individua nello spettro i valori delle ampiezze della fondamentale e delle armoniche dei suoni.

Anche dall'analisi del **cepstrum** si possono ricavare feature interessanti. Il cepstrum è il risultato della trasformata di Fourier applicata allo spettro in decibel di un segnale. Il nome deriva dal capovolgimento delle prime quattro

lettere della parola “spectrum”. E' molto efficace per discriminare la voce umana. Per le applicazioni musicali lo spettro viene trasformato secondo le bande di frequenza della scala mel, ossia una scala di percezione dell'altezza (“*pitch*”) del suono. Il risultato sono i coefficienti spettrali **Mel** o **MFCCs**, utilizzati in particolare per il riconoscimento vocale e alcune tecniche di *pitch-detection*. [038]

Sia nel dominio temporale che in quello delle frequenze esistono svariate feature che sono varianti o combinazioni di quelle sopra descritte. Esistono anche feature ricavabili da una rappresentazione ibrida chiamata **spettrogramma**. Lo spettrogramma vuole essere un superamento sia dei limiti del dominio del tempo che non visualizza le frequenze sia dello spettro sonoro che non mostra i picchi di frequenze. [034]

Generalmente le feature di basso livello non sono estrapolate dall'intero file audio, ma da **brevi frammenti**, talvolta di alcuni secondi; questo per evitare che i valori medi non perdano di significato perchè troppo approssimativi per elementi così variabili come la quantità di energia e le forme d'onda dei segnali. A volte si considerano pochi secondi come esemplificativi di tutto brano, scelti nella parte centrale dove solitamente si sviluppa il ritornello o comunque il picco espressivo della composizione [039], altre volte il pezzo viene analizzato nella sua totalità frammentandolo. Tuttavia non esistono regole definite circa la lunghezza dei frammenti e la loro scelta dipende dai fini dell'esperimento. Come detto la maggioranza dei file audio da cui vengono estrapolate tali feature sono WAV, MP3 o simili.

- Le **feature di alto livello** (“*high level features*”) rappresentano astrazioni musicali ricavabili da ascoltatori musicalmente allenati. Queste astrazioni riguardano gli **strumenti presenti** nel brano, il **profilo melodico**, la frequenza di certi **intervalli** o **accordi**, la **tonalità** predominante e le **proprietà ritmiche**. Molte feature di alto livello non possono essere estrapolate da registrazioni audio, al contrario sono ricavabili in modo relativamente semplice da file in

formati simbolici quali il MIDI, che memorizza eventi e parametri piuttosto che i *sample* reali. Le feature di alto livello riescono a rappresentare aspetti teorici irrecuperevoli con quelle di basso livello. Ecco allora che è utile implementare efficienti sistemi che trasformano i “*wave format*” in “*symbolic format*” per poi estrarre più comodamente le feature di questo tipo. Inoltre i formati simbolici consentono di trasformare facilmente in file musicali anche brani di cui si possiede solo lo spartito e non la registrazione, permettendo un'analisi di alto livello della **musica scritta**.

- Le **feature culturali** (“*cultural features*”) sono informazioni socio-culturali e commerciali che **trascendono il contenuto musicale** e, tra gli altri obiettivi, mirano ad individuare le caratteristiche del **pubblico destinatario** del brano in questione o altre proprietà riguardanti il **genere musicale**. Sono di solito ottenute attraverso ricerche automatiche sul web che forniscono dati statistici attraverso l'**analisi di playlist** o **siti commerciali** che memorizzano i trend di acquisto musicale o le **correlazioni tra le vendite** di prodotti diversi.

E' auspicabile la costruzione di sistemi di classificazione che si avvalgano di metodi che utilizzano feature appartenenti a tutte e tre le categorie descritte sopra.

I maggiori ostacoli alla realizzazione di questi sistemi sono la tendenza di ogni gruppo di ricerca a sviluppare solo le proprie tecnologie e la mancanza di standard diffusamente accettati in particolare circa il formato da utilizzare per memorizzare i valori delle feature. Nonostante ciò esistono degli standard de facto quali il formato “*Weka ARFF format*” (Witten and Frank 2005). Esiste anche un software di estrazione di feature, ritenuto da molti uno dei principali punti di riferimento sia accademici che industriali, chiamato **MARSYAS** (*Music Analysis, Retrieval and Synthesis for Audio Signals*, Tzanetakis and Cook 2000) . [040]

Il software appena citato, così come la maggioranza delle ricerche svolte fino ad oggi si basano prevalentemente su feature di basso livello. Lo sfruttamento del MIDI e delle feature di alto livello sono in una fase molto giovane di ricerca, per non parlare delle feature culturali la cui applicazione nei sistemi di classificazione è ancora più

recente. Nonostante ciò le feature di alto livello stanno portando a risultati molto soddisfacenti.

4.0

MIDI

Questo capitolo in primo luogo definisce il MIDI, raccontandone brevemente la storia e le esigenze che hanno spinto alla sua creazione, per poi analizzare le differenze tra un generico file WAV ed un generico file MIDI al fine di comprendere meglio la natura di entrambi.

In seguito si descrivono l'interfaccia ed il protocollo MIDI studiando la struttura di alcuni messaggi fondamentali e la funzione dei byte che li compongono.

Infine si analizza il contenuto dei midifile soffermandosi sulla struttura delle tracce, sugli eventi e sui metaeventi; i fondamentali elementi dei tre formati MIDI (formato 0, formato 1 e formato 2).

4.1

Introduzione al MIDI e alla Rappresentazione Simbolica dell'Informazione Musicale

L'acronimo **MIDI** (*Musical Instrument Digital Interface*) indica il protocollo standard di interazione tra strumenti musicali elettronici. E' riferito sia al linguaggio informatico che costituisce il protocollo di trasmissione sia all'interfaccia hardware che permette di veicolare tale trasmissione. Nasce agli inizi degli anni ottanta quindi potrebbe essere sostituito da protocolli moderni più performanti, tuttavia il suo successo è stato talmente rivoluzionario che ancora oggi sopravvive come standard indiscusso e gode di grande prestigio per la sua qualità e praticità, il basso costo, la leggerezza dei suoi file e l'enorme quantità di file già disponibili realizzati nel corso di

quasi trent' anni. La sua nascita risponde alla necessità di far comunicare strumenti musicali elettronici di marche differenti, che già hanno studiato metodi di interfacciamento proprietari, come Roland o Oberheim, superando i problemi riguardanti il *Timing* per le sincronizzazioni.

Il MIDI è presentato al pubblico nel 1983, alla fiera del *National Association of Music Merchants - NAMM* di Los Angeles. Ha un tale successo da creare divisioni nel comitato fondatore che si scinde in due gruppi; i produttori americani si uniscono nella *MMA (MIDI Manufactures Association)* mentre quelli giapponesi fondano la *JMA (Japan MIDI Association)*. Esistono alcune differenze tra le due case produttrici, risolte nel 1985 con la pubblicazione delle specifiche della versione 1.0 del protocollo da parte della *IMA (International MIDI Association)* che sancisce la istituzione di fatto del protocollo standard. [041]

Nel 1991 alla MMA e alla JSMC si ratifica una standardizzazione ancora più generale e potente chiamata **General MIDI (GM1)**. Il GM1 si prefigge di far suonare tutti i synth e i moduli sonori allo stesso modo nel senso che una base MIDI su un floppy una volta inserita in un qualsiasi player deve rispettare una serie di specifiche in modo da essere compatibile con tutti i modelli delle diverse marche.



Figura: il marchio di General MIDI

Nel 1999 nasce un nuovo accordo tra produttori di hardware e software che dà vita al **General MIDI Level 2 (GM2)**, un potenziamento del GM1. Le modifiche di questo standard si protraggono negli anni successivi, l'ultima risale al 2003.

Come detto i **dati** MIDI sono definibili **simbolici**. Sono simili a quello che un musicista legge o scrive su una partitura, ovvero un insieme di note che se eseguite producono suoni. Una stessa partitura MIDI può essere eseguita sfruttando diverse combinazioni di strumenti e può essere modificata prima, dopo o durante l'esecuzione. Un file WAV, a differenza del MIDI, non può essere facilmente modificato e non è possibile ricavare direttamente informazioni musicali da esso. Inoltre per quanto riguarda la memoria occupata, facendo un esempio pratico, tre minuti di registrazione audio su un file WAV pesano mediamente 30 MB, lo stesso brano in MIDI è codificato in circa 30 KB, ossia mille volte meno. Anzi va detto che il file MIDI corrispondente potrebbe essere addirittura vuoto, se quei tre minuti sono di silenzio. Nel **file WAV** sono presenti tutte le **informazioni relative al campionamento**, nel **file MIDI** invece sono presenti le **informazioni relative alle note suonate** e nessuna informazione circa i timbri ma solo un'indicazione del tipo di strumento da usare per riprodurre una certa traccia del brano. Per le parti cantate, infatti, esiste il grande problema di trovare strumenti sintetizzati che emulino la voce in modo sufficientemente credibile. Talvolta si impiegano dei *plug-in* che sono dei veri e propri cantanti virtuali come *Zero-G Leon & Lola*. [042]

4.2

Interfaccia e Protocollo MIDI

Il protocollo MIDI si basa su una **semplice trasmissione seriale asincrona**, in cui i singoli impulsi viaggiano attraverso il cavo MIDI ad una velocità prefissata di 31.250 bit/s. Essa è schematizzabile come segue:

- un bit di **start** per indicare l'inizio della trasmissione;
- otto bit che costituiscono il vero e proprio **messaggio**;
- un bit di **stop** per indicare la fine della trasmissione.

Ad ogni impulso di 0 Volt il bit corrisponde allo 0 digitale, e ad ogni impulso di +5 Volt il bit corrisponde ad 1. Il collegamento elettrico dell'interfaccia MIDI utilizza una presa a 5 poli di tipo DIN (*Deutsche Industrie Normen*) di cui sono impiegati solo tre connettori. In realtà basta solamente un cavo per trasmettere i segnali, gli altri due servono a chiudere il circuito di tensione a +5 Volt e a scaricare a massa eventuali disturbi presenti sulla linea.

Il flusso di dati di trasmissione è unidirezionale, quindi ogni strumento che fa uso di questo protocollo necessita di una porta **MIDI OUT** per inviare le informazioni ed una porta **MIDI IN** per riceverle. Opzionalmente è presente una porta **MIDI THRU** permette la ritrasmissione di dati ricevuti alla porta IN, consentendo il collegamento in cascata di diversi apparecchi.



Figura: le tre porte MIDI IN, THRU e OUT

La logica di lettura, scrittura e trasmissione si basa su due tipi di messaggi: lo **Status Byte** e il **Data Byte**. Il primo indica il tipo di messaggio ed il secondo le informazioni relative al messaggio. Se una tastiera A è collegata via MIDI ad un'altra tastiera B può inviarle il messaggio di riprodurre una determinata nota ad una certa intensità. Ciò non implica la trasmissione di nessun dato audio, ma solo le informazioni appena descritte, infatti la tastiera B può suonare la stessa nota usando un'altro timbro. E' per questa ragione che con il MIDI si diffondono ampiamente gli *expander*, ossia moduli sonori hardware senza tastiera, o più recentemente i *Virtual Synth*.

Dunque un **messaggio di nota** è composto da tre byte: lo **Status Byte** di Note On, il **Data Byte** che specifica l'**altezza** della nota e il **Data Byte** relativo all'**intensità** della

nota. Per prevenire errori lo Status Byte deve avere sempre il primo bit a 1, cioè impiega valori tra 128 e 255, mentre i Data Byte hanno il primo bit sempre a 0 e quindi utilizzano valori tra 0 e 127. Per fare un esempio se la tastiera A suona forte il DO centrale e lo trasmette alla tastiera B il messaggio sarà il seguente:

- Status Byte, Note On =144;
- Data Byte, Note = 60 (corrispondente al DO centrale);
- Data Byte, Velocity = 100 (numero che corrisponde ad una intensità forte).

Finchè si sta suonando il DO centrale sulla tastiera A il messaggio rimarrà valido per la tastiera B, fino a quando il dito non si alza dal tasto e parte il comando MIDI di fine esecuzione della nota:

- Status Byte, Note Off = 128;
- Data Byte, Note = 60;
- Data Byte, Velocity = 100;

oppure equivalentemente il messaggio che impone una intensità nulla alla nota:

- Status Byte, Note On = 144;
- Data Byte, Note = 60;
- Data Byte, Velocity = 0 .

La trasmissione sopra descritta riguarda un canale, ma per ogni porta MIDI si possono utilizzare fino a 16 canali. Il canale è specificato dallo Status Byte di Note On attraverso un numero compreso tra 144 (canale 1) e 159 (canale 16).

Nel protocollo MIDI si applica il concetto di **Running Status** per cui se lo Status Byte durante nelle istruzioni successive rimane lo stesso è inutile ritrasmetterlo perchè vale quello corrente.

Avendo a disposizione valori da 128 a 255 i Status Byte sono classificabili in due categorie in base al tipo di messaggio.

I **messaggi di canale** possono presentare i seguenti Status Byte:

- **Note On** che è già stato descritto;
- **Note Off** (128-143), rappresenta il messaggio di spegnimento della nota

- suonata ed è seguito da due Data Byte (Key Number e Release Velocity);
- **Polyphonic Aftertouch** (160-175), rappresenta un messaggio di effetto, ad esempio il Modulation, trasmesso dopo il Note On ed è seguito da un Data Byte Key Number che seleziona la nota e un Data Byte Aftertouch che indica quanto profondo debba essere l'effetto;
 - **Control Change** (176-191), rappresenta il messaggio trasmesso tipicamente dai *controller*, ossia dal pedale di Sustain, dagli slider delle Master Keyboard, dalla rotella del Modulation Wheel o dalle consolle MIDI, per scegliere o cambiare il tipo di controllo durante l'esecuzione su una tastiera o per gestire moduli sonori o Virtual synth in un Sequencer, ed è seguito da un Data Byte che identifica il numero di controllo (Control Number) e un Data Byte con i valori del parametro relativo al controllo (Control Data);
 - **Program Change** (192-207), rappresenta il messaggio che serve a cambiare quello che viene chiamato timbro, Patch o Preset ed è seguito da un solo Data Byte pertanto permette la selezione di 128 elementi differenti, se i timbri di un apparecchio MIDI sono in numero maggiore è perchè sono organizzati in diversi banchi a loro volta selezionabili con messaggi di Control Change;
 - **Channel Aftertouch** (208-223) rappresenta un tipo di messaggio particolare e poco usato perchè riguarda tutto il canale e non una singola nota suonata ed ha bisogno quindi di un solo Data Byte;
 - **Pitch Bend** (224-239) rappresenta il messaggio spedito dalla rotella o dal joystick di Pitch Bend, presente su tutti i sintetizzatori, che provoca una variazione del tono delle note suonate, ed è seguito da due Data Byte (Bend MSB e Bend LSB) che insieme indicano il valore proporzionale alle diverse possibili impostazioni di intonazione massima e minima.

Esistono poi i **messaggi di sistema** che non hanno canalizzazione MIDI e possono essere inviati in qualunque momento ed in qualunque canale con massima priorità di esecuzione e presentano i seguenti Status Byte:

- System Exclusive Start (240);

- System Exclusive End (247);
- MTC Quarter Frame (241);
- Song Pointer (242);
- Song Select (243);
- Tune Request (246);
- Timing Clock (248);
- Start (250);
- Continue (251);
- Stop (252);
- Active Sensing (254);
- Reset (255).

Tra tali messaggi sono significativi quelli di sistema esclusivo chiamati **SysEx**. Con questi comandi è possibile modificare radicalmente le impostazioni di sistema come gli **effetti interni**, l'**intonazione di base**, il **volume generale**, vari **parametri** relativi ai **timbri** e talvolta al **display** grafico e testuale dell'apparecchio. Si chiamano esclusivi perchè ogni costruttore che è dotato del proprio Manufacturer ID, rilasciato dagli enti MMA o JMSC, li implementa a modo proprio.

Esistono anche altri messaggi di sistema non esclusivi, come visto precedentemente; essi sono usati internamente per esempio da Drum Machine, console di controllo, Master KeyBoard, Sequencer e registratori multitraccia. [042]

4.3

Conformazione dei File MIDI

Le istruzioni ed i comandi del protocollo MIDI visti nel precedente paragrafo costituiscono il cuore dei cosiddetti **Standard MIDI File**. Il Midifile è nato come

formato di interscambio di song tra Sequencer hardware e software. Ha avuto un'enorme diffusione nel settore dell'intrattenimento dal vivo con l'impiego delle cosiddette basi MIDI soprattutto in seguito alla nascita del General MIDI. Anche nel settore delle suonerie dei cellulari sia mono che polifoniche il MIDI riscuote grande successo. Oltre che nel settore commerciale è usatissimo anche in campo didattico e di ricerca per la sua praticità e la sua rappresentazione simbolica dei dati musicali. Le informazioni trasmesse ad un apparecchio via MIDI costituiscono delle **tracce** e sono visualizzabili tramite un Sequencer sotto forma di lista ordinata. Oltre al listato è possibile visualizzare gli **eventi delle note** sotto forma di notazione musicale, rappresentazione più comoda per chi ha dimestichezza con gli spartiti.

Il midifile è composto da complesse strutture chiamate **Track Chunk** (Blocco-Traccia) formate da un certo numero di informazioni ed istruzioni. Ognuno di questi blocchi è fondamentale per la costituzione del midifile e per la sua esecuzione da parte del Sequencer.

La prima struttura, posta solo all'inizio e costituita da 14 byte, **Midi Track Header Chunk (MThd)**, ossia il blocco di intestazione specifica le informazioni di base circa il contenuto del file. Aperto il midifile con un lettore esadecimale il blocco di intestazione si presenta come segue dove **n** sono i byte variabili:

54 4D 68 64 00 00 00 06 00 0n 00 nn 0n nn .

- **54 4D 68 64** = definiscono il blocco come chunk di tipo *Mthd* .
- **00 00 00 06** = definiscono il numero di byte della restante parte del blocco *header*.
- **00 0n** = indicano il tipo di midifile, per il formato 0 $n = 0$, per il formato 1 $n = 1$ e per il formato 2 $n = 2$.

Ecco di seguito una descrizione dei tre formati.

Nel **formato 0** le tracce sono fuse in una singola che contiene però tutte le informazioni delle tracce originarie del brano. E' il più usato nelle tastiere, negli arranger e per la costruzione di basi Karaoke.

Nel **formato 1** le tracce sono salvate separatamente condividendo gli stessi valori di tempo e metrica, specificati nella prima traccia. Permette una gestione multitraccia del file ed è usatissimo nei Sequencer e riproduttori di midifile.

Il **formato 2** gestisce le tracce indipendentemente anche per quel che concerne il tempo e la metrica, ma è di scarsa diffusione soprattutto a causa della sua complessità.

- **00 nn** = indicano quante tracce sono presenti nel file MIDI.
- **0n nn** = indicano il “numero di impulsi per nota da $\frac{1}{4}$ ”: **PPQN** (*Pulse Per Quarter Note*). Ogni impulso è chiamato “*Tick*”. La durata in microsecondi di un Tick si ottiene facendo: $T = (60.000.000 / \text{bpm}) / \text{PPQN}$, dove **bpm** rappresenta i "battiti per minuto" (tempo metronomico) .

In seguito al *Mthd* seguono una o più tracce MIDI che contengono gli eventi descritti nel paragrafo precedente. Ogni blocco-traccia o *Midi Track* (**MTrk**) è preceduta dai seguenti 4 byte: **4D 54 72 6B** . Seguono altri 4 byte (**00 00 00 nn**) che indicano la lunghezza in byte della traccia. I primi byte danno informazioni circa il tempo, solitamente solo nella prima traccia chiamata infatti "Traccia del Tempo", i byte successivi rappresentano invece gli eventi musicali che possono essere "Eventi Midi", "Eventi SysEx" o "Eventi Dati Meta".

Le istruzioni di due qualsiasi degli eventi sono separate da un valore espresso in byte che indica la durata in PPQN dell'attesa tra l'esecuzione dei due eventi. Questo tempo di attesa è chiamato **Tempo Delta** ($T\Delta$, *Delta Time*) . [041]

Oltre agli eventi nelle tracce possono essere presenti i cosiddetti **Metaeventi** o **Metadati**. Essi sono dei messaggi che non hanno canale e non viaggiano sul bus MIDI, ma vengono solo letti o scritti dal software che importa o esporta il file. Presentano uno Status Byte sempre a 255 (FF in esadecimale), seguito prima da un byte che identifica il tipo di Metaevento, poi da uno che indica la lunghezza in byte dei dati del messaggio ed infine i dati veri e propri. Alcuni esempi di Metaeventi

sono: il **Sequence Number** (tipo 0) che indica il numero di sequenza, il **Text Event** (tipo 1) che serve per scrivere nel file brevi indicazioni in forma testuale, il **Copyright** (tipo 2) che serve per scrivere il copyright del brano, il **Track Name** (tipo 3), l' **Instrument Name** (tipo 4) usato per indicare il nome dello strumento usato nella traccia, il **Lyric** (tipo 5) usato nelle basi Karaoke per scrivere le sillabe del testo della canzone, il **Marker** (tipo 6) usato per indicare le posizioni delle parti che costituiscono la canzone quali l'Intro, le strofe e il ritornello, il **Cue Point** (tipo 7) simile al precedente, il **MIDI Channel Prefix** (tipo 32) usato nel formato 1 per indicare il canale della traccia, il **MIDI Port** (tipo 33) che indica la porta usata nella traccia, l'**End Of Track** (tipo 47) che è l'unico metaevento obbligatorio, il **Set Tempo** (tipo 81) usato per impostare il bpm, il **SMPTE Offset** (tipo 84) usato nella sincronizzazione di film indica il punto di inizio della traccia, il **Time Signature** (tipo 88) che indica il metro ritmico del brano, il **Key Signature** (tipo 89) che indica la tonalità del brano ed infine il **Sequencer Specific** (tipo 127) usato dal sequencer per scrivere dati specifici interni. [042]

PARTE SECONDA

5.0

Software Utilizzati per gli Esperimenti e Procedura di Esecuzione

Questo capitolo presenta i due software impiegati per gli esperimenti di questa tesi; JSymbolic che estrapola feature da file MIDI e JavaNNS che simula reti neurali, e i due eseguibili implementati appositamente per interfacciarli, JFeatureXmlExtractor e JPatConverter. Si descrivono i quattro programmi in ordine di esecuzione nel corso **di ogni esperimento** per chiarire nel dettaglio **la procedura utilizzata**. Essa consiste primariamente nell'estrazione dai file MIDI delle feature. Segue una operazione di normalizzazione dei valori delle feature tramite dei fogli di calcolo costruiti su misura. I valori normalizzati sono i pattern di ingresso delle reti neurali simulate da JavaNNS. Le reti devono essere addestrate per riconoscere, tramite l'analisi delle feature di questi brani musicali, il compositore tra due possibili.

5.1

JSymbolic

JSymbolic è un programma appartenente al pacchetto software open source JMIR, implementato in Java a fini di ricerca in ambito MIR. E' scaricabile da questo sito:

<http://sourceforge.net/projects/jmir>. Può essere utilizzato per studiare formati musicali di tipo "wave" o "symbolic" e per gestire collezioni di brani musicali. Il principale implementatore e designer è Cory McKay ed i finanziamenti al progetto provengono principalmente dalle società: "*Social Sciences and Humanities Research Council of Canada*", "*Canada Foundation for Innovation*" e "*Fonds Québécois de la recherche sur la société et la culture*". I software sono utilizzati per estrapolare feature di basso livello, alto livello e culturali, per applicare algoritmi di apprendimento a sistemi artificiali e per manipolare dati e metadati musicali.

In particolare JSymbolic si occupa di estrazione di feature di alto livello da file MIDI. Esso possiede una libreria di ben 160 feature. JSymbolic è concepito sia come piattaforma per implementare nuove feature, sia come tool che attraverso le feature estrapolate permette di classificare brani o più in generale trovare somiglianze tra file musicali.

Non è facile stabilire quante e quali feature sia necessario estrapolare per una classificazione ottimale. Ogni compositore presenta le sue feature discriminanti. Se le feature considerate sono poche si rischia di non cogliere le caratteristiche distintive dei brani in questione, al contrario se sono troppe è probabile che si fornisca un eccesso di informazioni che invece di aiutare gli strumenti di classificazione in un certo senso li "confondano". L'ideale è dunque un'applicazione che consideri un **gran numero di feature** sia generali che più specifiche e che dia la possibilità di **selezionarle in base alle esigenze di classificazione**. JSymbolic adotta proprio questa politica. Le **feature di alto livello** hanno il grande vantaggio che possono essere progettate sfruttando **conoscenze di teoria musicale e musicologia**, quindi abbracciano una gamma vastissima di aspetti e sono potenzialmente infinite. JSymbolic le classifica in due sottoclassi principali: quelle **unidimensionali** e quelle **multidimensionali**. Ognuna appartenente alla prima sottoclasse consiste in un singolo numero che rappresenta un unico aspetto. Le seconde invece sono costituite da diversi valori che considerati singolarmente hanno un significato limitato, ma valutati insieme possono rilevare pattern importanti. In questa tesi sono utilizzate solo

feature monodimensionali.

Le feature che offre JSymbolic possono essere suddivise in sette categorie in base ai diversi aspetti della composizione a cui si riferiscono.

- ***Instrumentation.*** Indagano quali strumenti sono presenti nel brano ed il loro livello di presenza/importanza. Valutano la rilevanza degli strumenti accordati (e.g. Piano, Chitarra, etc) e di quelli non accordati (e.g. Percussioni) e le relazioni tra loro intercorrenti.
- ***Texture.*** Studiano il numero di voci presenti, la loro rilevanza e come interagiscono (polifonia, omofonia).
- ***Rhythm.*** Misurano gli intervalli di tempo tra gli attacchi delle note e la loro durata. Individuano i metri ritmici, eventuali effetti ritmici utilizzati quali il "rubato" o la "sincope" e valutano la differenza ritmica tra le diverse voci.
- ***Dynamics.*** Misurano l'intensità delle note, le variazioni di intensità nel tempo e le differenze di volume da voce a voce.
- ***Pitch Statistics.*** Quantificano la ricorrenza assoluta delle note ("*pitch class*") e quella relativa alle note di ogni ottava ("*pitch*"). Valutano quanto sia tonale il brano e quanto vari la sua tonalità.
- ***Melody.*** Valutano che tipo di intervalli melodici presenta il brano e con quale frequenza. Individuano l'eventuale ripetizione di frasi melodiche e la loro variabilità.
- ***Chords.*** Valutano che tipo di intervalli verticali presenta il brano e con quale frequenza. Quantificano le variazioni armoniche del brano e la loro velocità di evoluzione.

[037]

Il codice di ogni feature di JSymbolic è in un modulo separato e per implementare una feature non è affatto necessario conoscere a fondo la struttura dell'intero programma. Il software è progettato per sviluppatori di diverso livello di conoscenze musicali ed informatiche per rispondere alle esigenze di una gamma di implementatori più vasta possibile.

L'interfaccia utente è molto intuitiva e semplice da usare. Tramite *check-box* è possibile selezionare le feature una ad una.

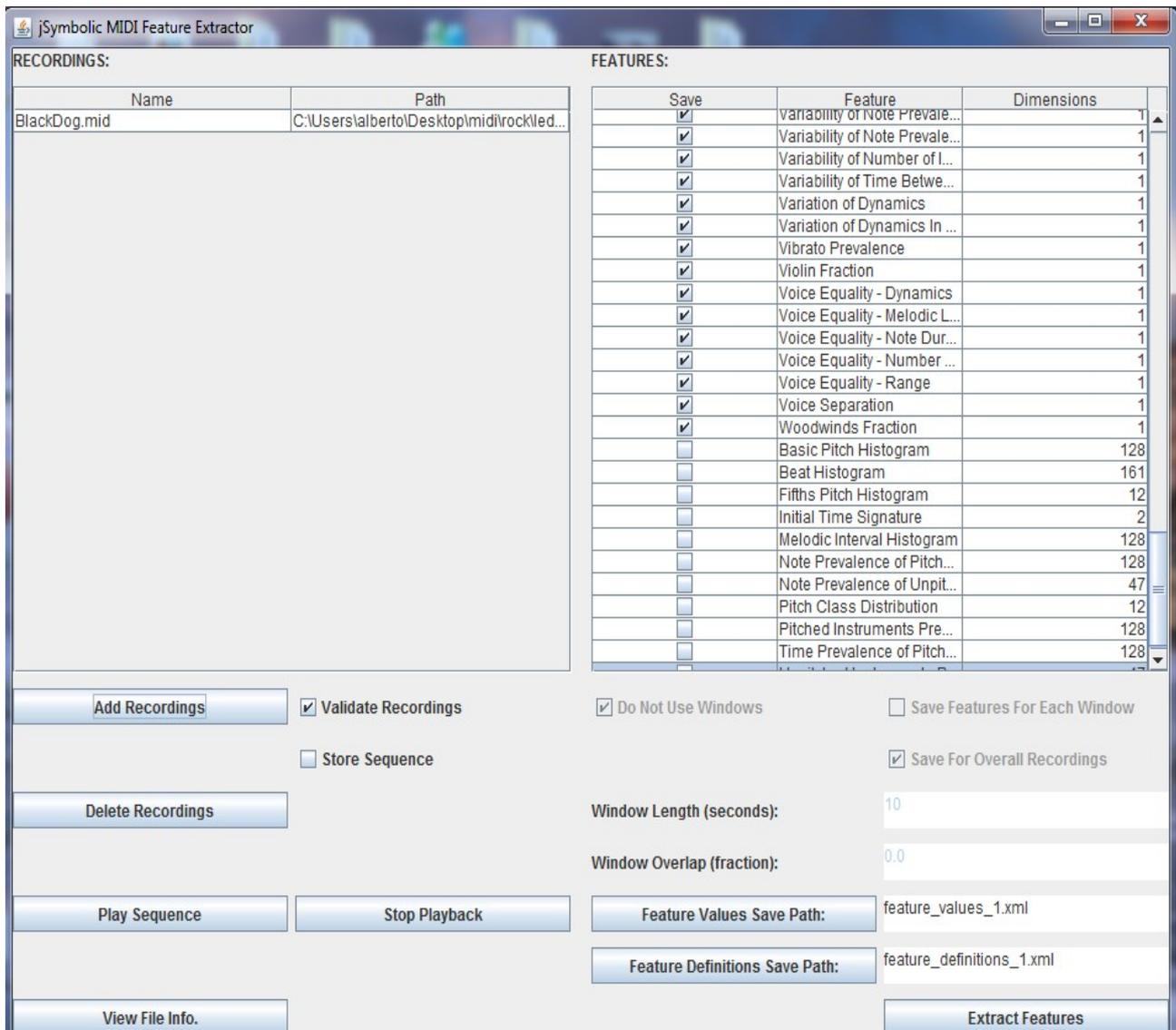


Figura: Interfaccia Utente di JSymbolic

Aggiunto alla lista il brano MIDI che si vuole analizzare e scelte le feature che si vogliono estrarre è sufficiente premere il pulsante *"Extract Features"* per ottenere il risultato desiderato.

Gli output del programma sono due file xml.

Il primo presenta una lista con la descrizione di ognuna delle feature selezionate, il secondo è l'elenco dei valori corrispondenti alle feature scelte scritte nella seguente notazione: "xEy", dove "x" e "y" sono due valori reali qualsiasi ed "E" è il numero di Nepero.

Il valore corrispondente si ottiene facendo il prodotto tra "x" ed "E" elevato alla (con esponente) "y". Es. $4,3E0 = 4,3$

```
<?xml version="1.0"?>
<!DOCTYPE feature_key_file>
- <feature_key_file>
  <comments/>
  - <feature>
    <name>Duration</name>
    <description>The total duration in seconds of the music.</description>
    <is_sequential>false</is_sequential>
    <parallel_dimensions>1</parallel_dimensions>
  </feature>
  - <feature>
    <name>Acoustic Guitar Fraction</name>
    <description>Fraction of all Note Ons belonging to acoustic guitar patches (General MIDI patches 25 to 26).</description>
    <is_sequential>true</is_sequential>
    <parallel_dimensions>1</parallel_dimensions>
  </feature>
  - <feature>
    <name>Amount of Arpeggiation</name>
    <description>Fraction of horizontal intervals that are repeated notes, minor thirds, major thirds, perfect fifths, minor sevenths,
      or major tenths.</description>
    <is_sequential>true</is_sequential>
    <parallel_dimensions>1</parallel_dimensions>
  </feature>
  - <feature>
    <name>Average Melodic Interval</name>
    <description>Average melodic interval (in semi-tones).</description>
    <is_sequential>true</is_sequential>
    <parallel_dimensions>1</parallel_dimensions>
  </feature>
  - <feature>
    <name>Average Note Duration</name>
    <description>Average duration of notes in seconds.</description>
    <is_sequential>true</is_sequential>
    <parallel_dimensions>1</parallel_dimensions>
  </feature>
  - <feature>
```

Figura: Frammento Iniziale di un File Xml Contenente
la Descrizione delle Feature Selezionate

```

<?xml version="1.0"?>
<!DOCTYPE feature_vector_file>
- <feature_vector_file>
  <comments/>
  - <data_set>
    <data_set_id>C:\Users\alberto\Desktop\midi\rock\ledZeppelin\BlackDog.mid</data_set_id>
    - <feature>
      <name>Duration</name>
      <v>3,044E2</v>
    </feature>
    - <feature>
      <name>Acoustic Guitar Fraction</name>
      <v>0E0</v>
    </feature>
    - <feature>
      <name>Amount of Arpeggiation</name>
      <v>4,438E-1</v>
    </feature>
    - <feature>
      <name>Average Melodic Interval</name>
      <v>3,675E0</v>
    </feature>
    - <feature>
      <name>Average Note Duration</name>
      <v>3,187E-1</v>
    </feature>
    - <feature>
      <name>Average Note To Note Dynamics Change</name>
      <v>1,095E-1</v>
    </feature>
    - <feature>
      <name>Average Number of Independent Voices</name>
      <v>3,172E0</v>
    </feature>
    - <feature>
      <name>Average Range of Glissandos</name>
      <v>0E0</v>
    </feature>
  </data_set>
</feature_vector_file>

```

Figura: Frammento Iniziale di un File Xml di Esempio Contenente
i Valori delle Feature Selezionate

5.2

JFeatureXmlExtractor

JFeatureXmlExtractor è un eseguibile in Java che è stato realizzato per estrapolare i valori delle feature dal file xml output di Jsymbolic.

Il programma ricavati i valori nella notazione "xEy" descritta nel precedente paragrafo dal file xml li converte in numeri reali con virgola rappresentati in Java con il tipo di dato *double*.

In seguito organizza i valori in una tabella in cui ogni colonna presenta in alto il nome di una feature e a seguire tutti i numeri corrispondenti. Ogni riga invece è riferita a un file MIDI diverso.

L'utente deve eseguire JFeatureXmlExtractor tante volte quanti sono i file xml corrispondenti ai diversi brani musicali scelti, digitando il seguente comando da console:

```
> java JFeatureXmlExtractor feature_values_1.xml Autor >> tabella_valori.ods
```

in cui "*feature_values_1.xml*" è il file xml contenente la lista dei valori delle feature estratte da un determinato file MIDI e "*Autor*" può essere sostituito con una qualsiasi stringa scelta dall'utente che identifica l'autore del brano presente nel file MIDI.

Viene poi utilizzato il comando " >> " per indirizzare in modalità append le righe della tabella dei valori su un foglio elettronico, in modo che ad ogni esecuzione la prima riga del file non venga sovrascritta.

Per gli esperimenti di questa tesi è stato utilizzato il programma di open office **Calc** quindi l'estensione è ".ods", ma è valido qualsiasi foglio di calcolo. L'estensione del file varia di conseguenza.

In fase di apertura del file "*tabella_valori.ods*" appare il menu di importazione testo in cui è necessario specificare come carattere di separazione tra i valori ed i nomi

delle feature il simbolo pipe " | ".

Le liste di feature nella tabella descritta sono riordinate secondo gli autori specificati in esecuzione ed organizzate in tanti insiemi quanti sono gli autori. In questa tesi gli autori dei brani analizzati sono sempre due per ogni esperimento.

5.3

Fogli di Calcolo

Costruiti per la Normalizzazione

Sono stati preparati dei modelli in Calc che normalizzano i valori della tabella output di JFeatureXmlExtractor. I modelli costruiti variano a seconda del numero di file MIDI che si è scelto di analizzare per l'esperimento, ma la loro struttura è sempre la stessa.

Una volta copiata all'inizio di uno di questi modelli la tabella output di JFeatureXmlExtractor, di seguito all'interno dello stesso file, compare una tabella identica che presenta gli stessi valori, ma normalizzati (ossia rappresentati in una scala tra 0 e 1). La normalizzazione è calcolata tramite la seguente formula:

$$x_{norm} = (x - x_{min}) / (x_{max} - x_{min})$$

dove " x_{norm} " è il nuovo valore normalizzato, " x " è il valore che deve essere convertito nella nuova rappresentazione normalizzata, " x_{min} " e " x_{max} " sono il minimo ed il massimo valore tra tutti quelli di una stessa colonna della tabella, ossia relativi alla stessa feature.

Nella nuova tabella di valori normalizzati risulterà quindi che il minimo sarà rappresentato come "0", il massimo come "1" e tutti gli altri come numeri con virgola intermedi tra "0" e "1".

	A	B	C	D	E	F	G	H	I
1		Amount of Arpeggiation	Average Melodic Interval	Chromatic Motion	Importance of High Register	Importance of Loudest Voice	Melodic Octave	Most Common Melodic Interval	Overall Dynamic Range
2	Bach	0,455803	2,323000	1,039995	0,929263	0,000000	0,141696	1,996850	5,436564
3	Bach	0,810438	2,991000	0,776961	0,910134	0,000000	0,185409	1,717629	5,980220
4	Bach	0,490751	2,798000	1,031534	0,996953	0,000000	0,182567	1,759567	0,000000
5	Bach	0,785055	3,078000	1,217313	0,903144	0,000000	0,195830	1,217313	5,980220
6	Bach	1,235707	2,588000	0,864885	1,091866	0,000000	0,185951	1,084141	13,047753
7	Bach	0,657401	2,409000	1,035213	0,907559	0,000000	0,189290	1,741541	8,426674
8	Bach	0,939932	2,811000	0,955383	1,060596	0,000000	0,205574	1,407507	0,000000
9	Bach	1,158452	3,080000	0,674691	0,838765	0,000000	0,468260	1,241225	5,708392
10	Bach	0,962005	2,552000	1,462689	0,640110	0,000000	0,517387	1,462689	5,980220
11	Bach	0,590447	2,587000	1,009829	0,766293	0,000000	0,399591	1,759567	0,000000
12	Bach	1,759567	3,584000	0,734287	0,774386	0,000000	0,560694	0,817428	5,436564
13	Bach	0,993642	3,223000	0,969730	0,963844	0,000000	0,263633	1,213634	5,436564
14	Bach	0,740173	2,762000	1,152934	0,953544	0,000000	0,353360	1,356371	5,980220
15	Bach	1,058389	3,149000	0,832879	0,531218	0,000000	0,340504	1,347910	5,980220
16	Mozart	1,635592	4,900000	0,708536	1,242697	3,533766	0,672210	0,708536	14,678722
17	Mozart	1,770236	5,174000	0,580882	1,008358	8,000000	0,684391	0,622820	14,406894
18	Mozart	1,303397	6,884000	0,571685	1,235707	0,000000	0,494786	0,743852	0,000000
19	Mozart	1,583721	4,229000	0,648939	1,310754	3,261938	0,735412	0,783951	26,910990
20	Mozart	1,554291	4,558000	0,726930	1,261826	5,000000	0,550273	0,726930	19,571629
21	Mozart	1,409714	4,641000	0,636064	1,615727	5,000000	0,559205	0,940300	10,329471
22	Mozart	1,253365	4,143000	0,470886	1,117618	3,000000	0,371360	1,393527	7,611189
23	Mozart	1,323630	2,862351	0,429683	1,033005	0,000000	0,536740	0,463528	21,474426
24	Mozart	1,659872	2,984673	0,801185	1,076783	0,000000	1,265791	0,430051	23,105396
25	Mozart	1,502788	9,300000	1,068607	1,530378	0,000000	1,010819	0,591918	23,105396
26	Mozart	1,794148	3,783000	0,517238	0,671748	0,000000	0,689398	0,784319	16,309691
27	Mozart	1,328781	4,257000	0,549612	1,200758	8,000000	0,451885	0,866356	7,611189
28	Mozart	1,756624	4,109000	0,485969	0,822946	0,000000	0,479250	0,808599	17,125176
29	Mozart	1,631177	4,183000	0,567270	1,412657	0,000000	0,539311	0,732816	17,397004
30									
31	VALORI NORMALIZZATI								
32		0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000

Figura: Esempio di Tabella dei Valori,
Frammento Iniziale del Foglio di Calcolo

	A	B	C	D	E	F	G	H	I
30									
31	VALORI NORMALIZZATI								
32	Bach	0,000000	0,000000	0,590812	0,367028	0,000000	0,000000	1,000000	0,202020
33	Bach	0,264981	0,095743	0,336182	0,349389	0,000000	0,038888	0,821789	0,222222
34	Bach	0,026113	0,068081	0,582621	0,429444	0,000000	0,036359	0,848556	0,000000
35	Bach	0,246014	0,108213	0,762464	0,342944	0,000000	0,048158	0,502465	0,222222
36	Bach	0,582738	0,037982	0,421296	0,516961	0,000000	0,039369	0,417469	0,484848
37	Bach	0,150632	0,012326	0,586182	0,347015	0,000000	0,042340	0,837051	0,313131
38	Bach	0,361737	0,069944	0,508903	0,488128	0,000000	0,056826	0,623855	0,000000
39	Bach	0,525014	0,108499	0,237179	0,283582	0,000000	0,290513	0,517727	0,212121
40	Bach	0,378230	0,032822	1,000000	0,100407	0,000000	0,334216	0,659075	0,222222
41	Bach	0,100605	0,037839	0,561610	0,216757	0,000000	0,229425	0,848556	0,000000
42	Bach	0,974162	0,180737	0,294872	0,224220	0,000000	0,372743	0,247241	0,202020
43	Bach	0,401869	0,128995	0,522792	0,398915	0,000000	0,108476	0,500117	0,202020
44	Bach	0,212479	0,062921	0,700142	0,389417	0,000000	0,188298	0,591219	0,222222
45	Bach	0,450247	0,118389	0,390313	0,000000	0,000000	0,176860	0,585818	0,222222
46	Mozart	0,881528	0,369356	0,269943	0,656038	0,441721	0,471948	0,177741	0,545455
47	Mozart	0,982133	0,408628	0,146368	0,439959	1,000000	0,482784	0,123034	0,535354
48	Mozart	0,633315	0,653719	0,137464	0,649593	0,000000	0,314110	0,200282	0,000000
49	Mozart	0,842771	0,273183	0,212251	0,718792	0,407742	0,528172	0,225875	1,000000
50	Mozart	0,820781	0,320338	0,287749	0,673677	0,625000	0,363472	0,189481	0,727273
51	Mozart	0,712754	0,332234	0,199786	1,000000	0,625000	0,371418	0,325663	0,383838
52	Mozart	0,595932	0,260857	0,039886	0,540706	0,375000	0,204310	0,614933	0,282828
53	Mozart	0,648433	0,077304	0,000000	0,462687	0,000000	0,351433	0,021367	0,797980
54	Mozart	0,899670	0,094836	0,359632	0,503053	0,000000	1,000000	0,000000	0,858586
55	Mozart	0,782298	1,000000	0,618510	0,921303	0,000000	0,773176	0,103311	0,858586
56	Mozart	1,000000	0,209259	0,084758	0,129579	0,000000	0,487238	0,226109	0,606061
57	Mozart	0,652281	0,277197	0,116097	0,617368	1,000000	0,275945	0,278469	0,282828
58	Mozart	0,971963	0,255984	0,054487	0,268996	0,000000	0,300290	0,241606	0,636364
59	Mozart	0,878230	0,266590	0,133191	0,812754	0,000000	0,353720	0,193238	0,646465
60									

Figura: Esempio di Tabella dei Valori Normalizzati,
Secondo Frammento del Foglio di Calcolo

5.4

JPatConverter

JPatConverter è un eseguibile in Java che è stato realizzato per creare i file di estensione ".pat" necessari a JavaNNS per la simulazione di apprendimento.

In ingresso l'eseguibile riceve la tabella dei valori normalizzati delle feature di tutti i brani dei due autori che si vogliono analizzare.

I file di output sono due. Uno costituisce il Training Set della rete, l'altro il Validation Set. La **struttura dei due file ".pat"** è la stessa. Inizialmente è specificato il **numero**

di pattern ossia il numero di esempi a cui verrà sottoposta la rete neuronale ai fini dell'apprendimento nel caso del Training Set e ai fini del test nel caso del Validation Set. Gli esempi sono tanti quanti sono i brani che si è scelto di analizzare. Segue il **numero delle unit di input** della rete che negli esperimenti di questa tesi è variabile e corrisponde al numero di feature che si è scelto di prendere in considerazione. Il **numero delle unit di output** corrisponde alle uscite della rete e negli esempi del presente elaborato è sempre uguale a uno. Basta soltanto una unit infatti per indicare se il brano in questione appartiene al primo autore (la unit assume valore 0) o al secondo (la unit assume valore 1). Segue l'elenco degli input di ogni pattern con il corrispondente output per consentire appunto l'apprendimento o il test di apprendimento alla rete neuronale.

```

SNNS pattern definition file V3.2
generated at 31-1-2012

No. of patterns : 20
No. of input units : 10
No. of output units : 1

# Input pattern 1:
  0.000000  0.000000  0.590812  0.367028  0.000000  0.000000  1.000000  0.202020  1.000000  0.478364
# Output pattern 1:
0
# Input pattern 2:
  0.264981  0.095743  0.336182  0.349389  0.000000  0.038888  0.821789  0.222222  0.787975  0.281724
# Output pattern 2:
0
# Input pattern 3:
  0.026113  0.068081  0.582621  0.429444  0.000000  0.036359  0.848556  0.000000  0.903769  0.320066
# Output pattern 3:
0
# Input pattern 4:
  0.246014  0.108213  0.762464  0.342944  0.000000  0.048158  0.502465  0.222222  0.759350  0.266204
# Output pattern 4:
0
# Input pattern 5:
  0.582738  0.037982  0.421296  0.516961  0.000000  0.039369  0.417469  0.484848  0.574655  0.752419
# Output pattern 5:
0

```

Figura: Esempio di Frammento iniziale di un file ".pat" di Training Set

```
No. of patterns : 8
No. of input units : 10
No. of output units : 1

# Input pattern 1:
  0.974162  0.180737  0.294872  0.224220  0.000000  0.372743  0.247241  0.202020  0.378740  0.722476
# Output pattern 1:
0
# Input pattern 2:
  0.401869  0.128995  0.522792  0.398915  0.000000  0.108476  0.500117  0.202020  0.666283  0.511594
# Output pattern 2:
0
# Input pattern 3:
  0.212479  0.062921  0.700142  0.389417  0.000000  0.188298  0.591219  0.222222  0.793585  0.300164
# Output pattern 3:
0
# Input pattern 4:
  0.450247  0.118389  0.390313  0.000000  0.000000  0.176860  0.585818  0.222222  0.665276  0.373380
# Output pattern 4:
0
# Input pattern 5:
  1.000000  0.209259  0.084758  0.129579  0.000000  0.487238  0.226109  0.606061  0.306530  1.000000
# Output pattern 5:
1
```

Figura: Esempio di Frammento iniziale di un file ".pat" di Validation Set

Dopo aver copiato in un file di testo la tabella dei valori normalizzati presente nel foglio di calcolo descritto nel paragrafo precedente è possibile eseguire JPatConverter con tale comando da console:

```
> java JPatConverter tabellaNorm.txt Train.pat Val.pat 10 4
```

dove " *tabellaNorm.txt* " è il file di testo in cui è stata copiata la tabella, " *Train.pat* " e " *Val.pat* " rappresentano i nomi dei due file, scelti dall'utente e corrispondenti al Training Set e al Validation Set, che sono l'output del programma. I due numeri "10" e "4" sono il numero di brani analizzati (pattern) per ogni autore presenti rispettivamente nel Training Set e nel Validation Set. Questi due numeri particolari ("10" e "4") si riferiscono agli esempi delle due figure più sopra in cui appunto nel Training Set i pattern totali sono 20 (10+10) e nel Validation Set sono 8 (4+4).

In realtà è a **discrezione dell'utente scegliere come distribuire nei due file gli esempi totali disponibili** e cambiare i numeri nel comando da console di

conseguenza.

5.5

JavaNNS

Java Neural Network Simulator (JavaNNS) è un simulatore in Java per reti neurali sviluppato al "Wilhelm-Schickard-Institute for Computer Science" (WSI) a Tubinga in Germania principalmente da Igor Fischer, Fabian Hennecke, Christian Bannes e Andreas Zell, terminato negli anni 2001/2002. E' open source e scaricabile dal seguente sito <http://www.ra.cs.uni-tuebingen.de/software/JavaNNS/>. JavaNNS è basato sul kernel Stuttgart Neural Network Simulator (SNNS) 4.2, ma presenta una nuova interfaccia grafica anch'essa scritta in Java.

Il programma permette di usare reti neurali predefinite e di crearne di nuove al fine di applicare su di esse algoritmi di apprendimento ed analizzarle.

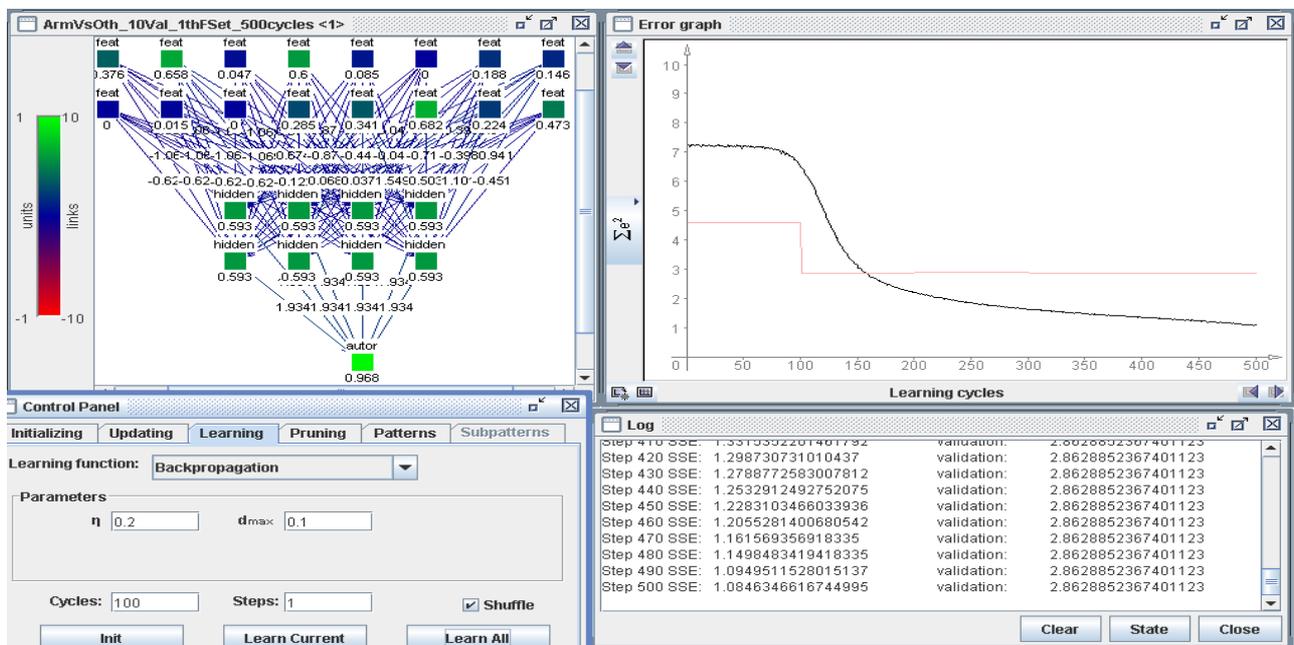


Figura: I Tool principali di Java NNS

Osservando la figura sopra si notano i quattro principali tool di Java NNS.

Quello in alto a sinistra è il **Graphical Network Display**. In esso si osservano le unit dei diversi strati della rete (input, hidden, output) e le connessioni tra di esse. Sotto ad ogni unit si legge il valore corrispondente alla unit stessa. Oltre all'indicatore numerico esiste anche una scala cromatica, come specificato nella legenda a sinistra della finestra, che nel caso di valori normalizzati va da un minimo di -1 corrispondente al rosso ad un massimo di +1 corrispondente al verde. Il blu corrisponde allo 0. Esistono ovviamente tutte le sfumature intermedie che variano a seconda del valore. Allo stesso modo anche le connessioni presentano un numero che indica il loro peso ("*weight*") e che ha una corrispondenza cromatica.

Non solo il *Graphical Network Display* visualizza graficamente le reti neurali salvate, ma è possibile organizzare manualmente le unit su di esso con il "drag and drop", nominarle e creare da nuovo tutti i livelli di una rete a proprio piacimento.

In basso a sinistra si trova il **Control Panel** che è la finestra principale di JavaNNS perchè è tramite esso che si fanno quasi tutte le modifiche e le modellazioni da apportare alla rete. Permette di inizializzare i pesi della rete, selezionare il Training e il Validation Set, impostare il numero di cicli e l'algoritmo di apprendimento e settarne i parametri realitivi.

Quello in alto a destra è l'**Error Graph** che di default presenta in ascissa il numero dei cicli di apprendimento ed in ordinata la somma degli errori al quadrato per ogni ciclo. Per errore in questo caso si intende la differenza tra il valore specificato negli output dei pattern presenti nei file ".pat" di Training o Validation e il valore reale di output ottenuto con gli stessi ingressi applicati alla rete neurale che si sta addestrando.

La **linea nera** del grafico si riferisce ai pattern appartenenti al **Training Set**.

La **linea rossa**, invece, si riferisce ai pattern appartenenti al **Validation Set**.

In basso a destra si trova il **Log** che tiene traccia testuale di tutti gli *step* di esecuzione del programma o di eventuali segnalazioni d'errore. In particolare durante i cicli di apprendimento mostra passo per passo il valore esatto della somma dei quadrati degli errori sia del Training che del Validation Set aiutando la lettura del grafico.

6.0

Esperimenti

Il capitolo descrive gli esperimenti svolti in questa tesi. Tutti rispettano la **procedura** descritta nel capitolo precedente. **L'obiettivo è addestrare una rete neurale al riconoscimento dell'autore di brani musicali mai analizzati.** Si considerano **due autori per volta** e li si confronta tramite determinate feature che saranno necessarie alla rete a discriminarli. Il numero totale di brani è suddiviso in due gruppi: uno costituirà il **training set** e l'altro il **validation set**.

Si considerano autori e band appartenenti a generi musicali completamente differenti. Si confrontano Bach e Mozart, Chet Baker e Luis Armstrong, Beatles e Rolling Stones e Beatles e Led Zeppelin. Per ogni coppia di autori si fanno diversi esperimenti per cercare le combinazioni di parametri e feature che portano al risultato di addestramento migliore.

6.1

Bach e Mozart

Si mettono a confronto Bach e Mozart analizzando delle composizioni per tastiera. Sono selezionate le prime 14 invenzioni a due voci di Bach e 14 movimenti casuali estrapolati dalle sonate per pianoforte di Mozart n.1, n.2, n.5, n.8, n.9, n.10, n.12 e n.16. Osservando la lista delle feature selezionabili di JSymbolic è possibile eliminarne alcune che è intuibile siano inutili o poco rilevanti: ad esempio quelle che riguardano strumenti musicali che non siano le tastiere.

Si estrapolano poi tutte le restanti feature. In questo modo si coniugano, nel processo di scelta delle più significative, criteri di teoria musicale con riscontri visivi dei valori presenti nella tabella nel foglio elettronico al fine di trovare quelle più significative. Si nota infatti in alcuni fortunati casi che i valori riguardanti ognuno dei due autori sono simili tra loro e visibilmente diversi al variare dell'autore a partire talvolta dalla prima cifra significativa.

Con tale procedimento si selezionano le 20 feature ritenute più distintive.

A loro volta queste 20 son suddivise in due set da 10, il primo con le feature giudicate "migliori" il secondo con le restanti.

	A	B	C	D	E	F	G	H	I	J	K	L
1		Amount of Arpeggiation	Average Melodic Interval	Chromatic Motion	Importance of High Register	Importance of Loudness	Melodic Octaves	Most Common Melody	Overall Dynamic Range	Stepwise Motion	Variability of Note Durations	
2	Bach	0,455803	2,323000	1,039995	0,929263	0,000000	0,141696	1,996850	5,436564	3,036845	1,394263	
3	Bach	0,810438	2,991000	0,776961	0,910134	0,000000	0,185409	1,717629	5,980220	2,494590	0,998057	
4	Bach	0,490751	2,798000	1,031534	0,996953	0,000000	0,182567	1,759567	0,000000	2,790733	1,075312	
5	Bach	0,785055	3,078000	1,217313	0,903144	0,000000	0,195830	1,217313	5,980220	2,421382	0,966787	
6	Bach	1,235707	2,588000	0,864885	1,091866	0,000000	0,185951	1,084141	13,047753	1,949025	1,946450	
7	Bach	0,657401	2,409000	1,035213	0,907559	0,000000	0,189290	1,741541	8,426674	2,776754	1,360050	
8	Bach	0,939932	2,811000	0,955383	1,060596	0,000000	0,205574	1,407507	0,000000	2,362890	1,105110	
9	Bach	1,158452	3,080000	0,674691	0,838765	0,000000	0,468260	1,241225	5,708392	1,915916	1,048089	
10	Bach	0,962005	2,552000	1,462689	0,640110	0,000000	0,517387	1,462689	5,980220	2,373926	1,225039	
11	Bach	0,590447	2,587000	1,009829	0,766293	0,000000	0,399591	1,759567	0,000000	2,769396	1,395735	
12	Bach	1,759567	3,584000	0,734287	0,774386	0,000000	0,560694	0,817428	5,436564	1,447973	1,886118	
13	Bach	0,993642	3,223000	0,969730	0,963844	0,000000	0,263633	1,213634	5,436564	2,183364	1,461217	
14	Bach	0,740173	2,762000	1,152934	0,953544	0,000000	0,353360	1,356371	5,980220	2,508938	1,035213	
15	Bach	1,058389	3,149000	0,832879	0,531218	0,000000	0,340504	1,347910	5,980220	2,180789	1,182732	
16	Mozart	1,635592	4,900000	0,708536	1,242697	3,533766	0,672210	0,708536	14,678722	1,288314	0,452860	
17	Mozart	1,770236	5,174000	0,580882	1,008358	8,000000	0,684391	0,622820	14,406894	1,203702	0,688670	
18	Mozart	1,303397	6,884000	0,571685	1,235707	0,000000	0,494786	0,743852	0,000000	1,315169	0,446974	
19	Mozart	1,583721	4,229000	0,648939	1,310754	3,261938	0,735412	0,783951	26,910990	1,433258	0,453227	
20	Mozart	1,554291	4,558000	0,726930	1,261826	5,000000	0,550273	0,726930	19,571629	1,226878	0,572788	
21	Mozart	1,409714	4,641000	0,636064	1,615727	5,000000	0,559205	0,940300	10,329471	1,576363	0,445134	
22	Mozart	1,253365	4,143000	0,470886	1,117618	3,000000	0,371360	1,393527	7,611189	1,864781	0,819635	
23	Mozart	1,323630	2,862351	0,429683	1,033005	0,000000	0,536740	0,463528	21,474426	0,808967	0,628338	
24	Mozart	1,659872	2,984673	0,801185	1,076783	0,000000	1,265791	0,430051	23,105396	0,479347	0,430419	
25	Mozart	1,502788	9,300000	1,068607	1,530378	0,000000	1,010819	0,591918	23,105396	0,882543	0,750842	
26	Mozart	1,794148	3,783000	0,517238	0,671748	0,000000	0,689398	0,784319	16,309691	1,263298	2,445295	
27	Mozart	1,328781	4,257000	0,549612	1,200758	8,000000	0,451885	0,866356	7,611189	1,415968	1,135644	
28	Mozart	1,756624	4,109000	0,485969	0,822946	0,000000	0,479250	0,808599	17,125176	1,125343	1,150727	
29	Mozart	1,631177	4,183000	0,567270	1,412657	0,000000	0,539311	0,732816	17,397004	1,300454	0,515031	

Figura: Bach vs Mozart, primo set di feature

Si ricorda una volta per tutte che nella descrizione degli esperimenti in questo capitolo sono mostrate esclusivamente le tabelle contenenti i valori non normalizzati estratti dai midifile, quindi non esattamente quelli che costituiscono i pattern usati per l'addestramento della rete. Questo perchè valori tutti compresi tra 0 e 1 sarebbero più difficili da leggere ed interpretare, poichè tutti addensati nello

stesso intervallo.

Come si evince dalla tabella le prime 10 feature scelte sono:

- "**Amount of Arpeggiation**", la quantità di intervalli orizzontali che sono note ripetute, terze minori, terze maggiori, quinte giuste, settime minori, settime maggiori, ottave, decime minori o decime maggiori;
- "**Average Melodic Interval**", la media degli intervalli melodici espressa in numero di semitoni;
- "**Chromatic Motion**", quantità di intervalli melodici corrispondenti ad uno spostamento di un semitono;
- "**Importance of High Register**", la quantità di note presenti nell'intervallo dei toni MIDI corrispondenti ai numeri dal 73 al 127;
- "**Importance of Loudest Voice**", la differenza tra il volume medio del canale con volume più alto e la media dei volumi degli altri canali che hanno almeno una nota;
- "**Melodic Octaves**", la quantità degli intervalli di ottava;
- "**Most Common Melodic Interval Prevalence**", la quantità di intervalli melodici che appartengono all'intervallo più comune;
- "**Overall Dynamic Range**", la differenza tra il picco di volume più alto ed il picco di volume più basso;
- "**Stepwise Motion**", la quantità di intervalli melodici che sono una seconda minore o maggiore;
- "**Variability of Note Duration**", la deviazione standard della durata delle note in secondi.

Bach e Mozart hanno due stili compositivi facilmente riconoscibili dall'orecchio umano, ma poiché questi due stili sottendono pattern molto definiti che rispondono spesso a **regole matematiche precise** anche il calcolatore risulta facilitato nell'individuazione delle caratteristiche distintive.

Molte differenze nei valori delle feature infatti sono individuabili anche dopo un'occhiata veloce e superficiale alla tabella riportata sopra.

In particolare va sottolineato che i brani di **Bach**, da molti chiamato "*il re del contrappunto*", sono caratterizzati da **due voci** (invenzioni a due voci appunto) **dotate di stessa importanza sia armonico-melodica che dinamica**.

Mozart invece, seppur le voci delle sonate siano magistralmente calibrate, conferisce alla **linea nella chiave di basso** una **funzione** perlopiù **armonica e di accompagnamento** alla **linea sulla chiave di violino** che presenta i **temi principali**.

Queste considerazioni teoriche hanno trovato riscontro durante la fase di analisi delle feature. Molte feature riguardanti l'importanza delle voci coglievano questo concetto. Se ne è scelta solo una per tutte, ossia "Importance of Loudest Voice", per evitare di costruire un set con informazioni ridondanti.

Gran parte delle 10 feature elencate riguardano gli intervalli melodici utilizzati dai due compositori, gli aspetti ritmici sono praticamente assenti tra le caratteristiche giudicate distintive.

Un'altra osservazione rilevante riguarda le dinamiche, quindi le variazioni di volume durante i brani. La feature "Overall Dynamic Range" evidenzia l'utilizzo molto più accentuato delle dinamiche in Mozart. Le composizioni di Bach presentano una forte omogeneità di volume sia temporale, sia tra le diverse voci.

Si procede con la costruzione tramite JavaNNS della rete neuronale che deve essere addestrata. Gli ingressi sono 10 come il numero di feature scelte. Si costruisce poi uno strato hidden con 5 units e uno strato di output con una sola unit che presenta in uscita il valore relativo all'autore. Si è scelto lo "0" per rappresentare Bach e l'"1" per rappresentare Mozart. La rete è di tipo Feedforward così come tutte le reti degli esperimenti di questa tesi, quindi tutte le unit di ogni strato sono collegate con tutte le unit dello strato successivo e precedente.

Dei 14 brani analizzati di ognuno dei due compositori, 10 fanno parte del training set e 4 del validation set. Quindi il training set presenta 20 pattern ed il validation set 8 pattern.

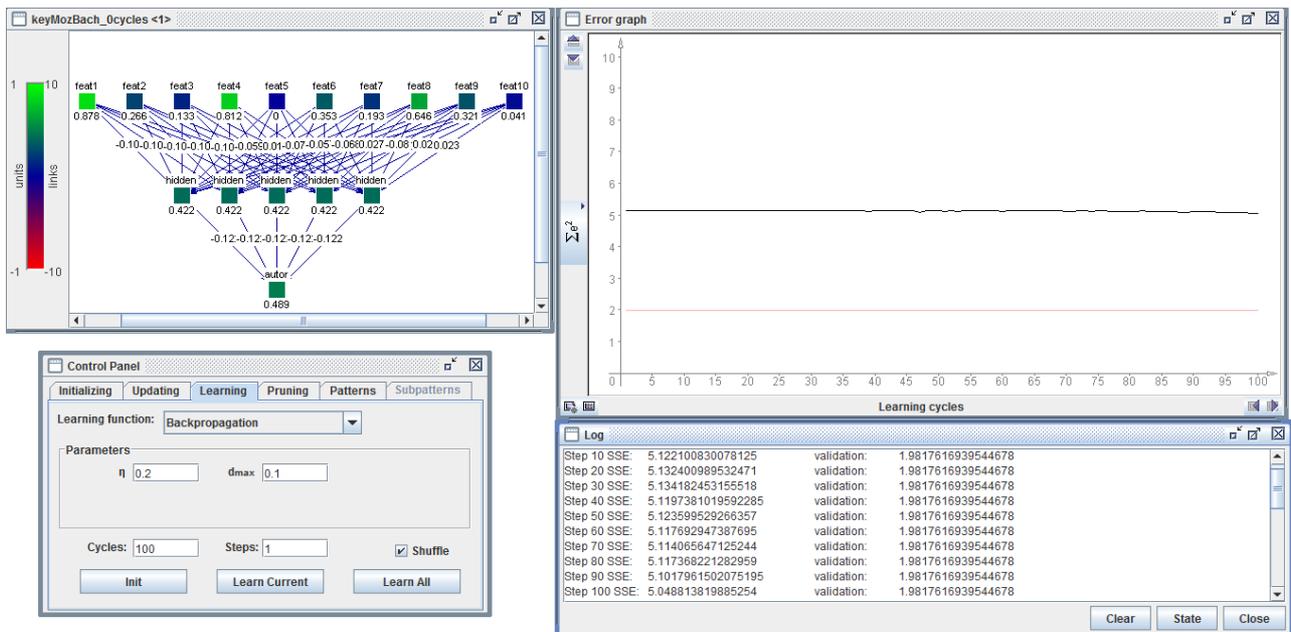


Figura: Bach vs Mozart, primo feature set,
rete neurale dopo 100 cicli di apprendimento

La figura mostra la struttura della rete descritta e le impostazioni principali. L'algoritmo di apprendimento è quello di Backpropagation e sarà lo stesso per tutti gli esperimenti di questa tesi. Si seleziona la modalità *shuffle* nel pannello di controllo che garantisce una scelta casuale dei pattern del training e del validation set durante l'addestramento. I parametri η e d_{max} di default sono rispettivamente a 0,2 e a 0,1 e sono sempre lasciati a questi valori se non esplicitamente specificato. Il numero di cicli di apprendimento impostato è 100 per ogni passo.

Dopo questa fase di settaggio si preme il pulsante *Learn All* del pannello di controllo che esegue i primi 100 *learning cycle*. I risultati sono visibili sull'*Error Graph* in forma grafica e sul *Log* in forma numerica precisa.

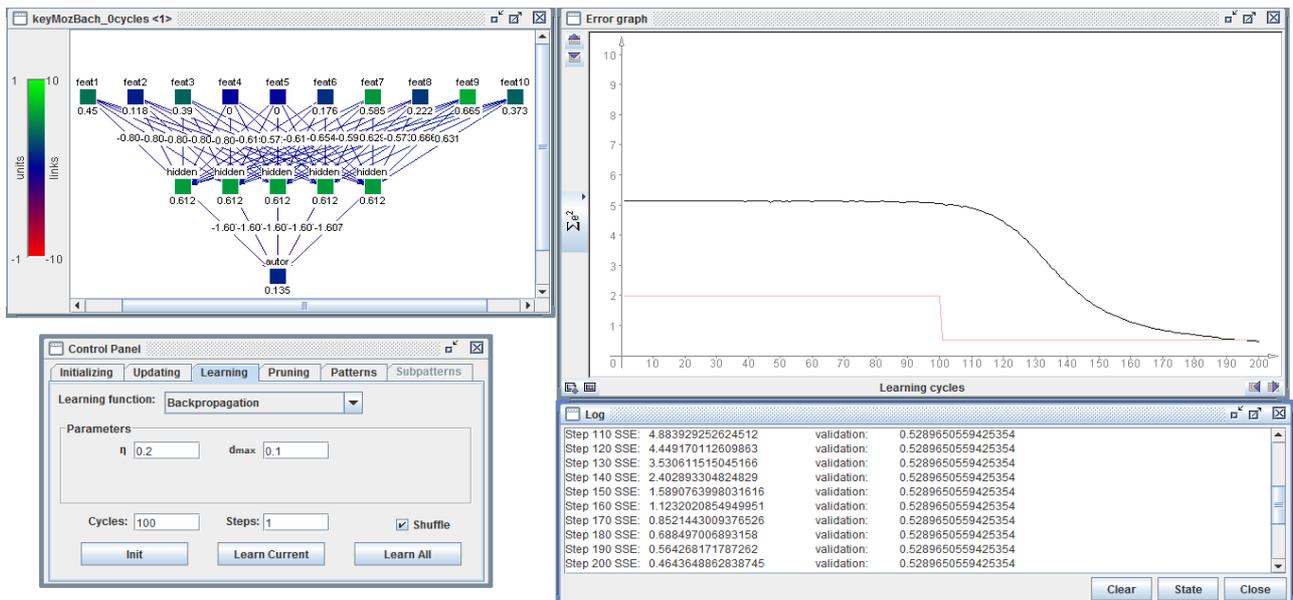


Figura: Bach vs Mozart, primo feature set, rete neuronale dopo 200 cicli di apprendimento

Si nota che dopo 100 cicli la rete comincia a migliorare le sue capacità di generalizzazione poichè la linea rossa che rappresenta la somma degli errori al quadrato riferiti al validation set scende sensibilmente.

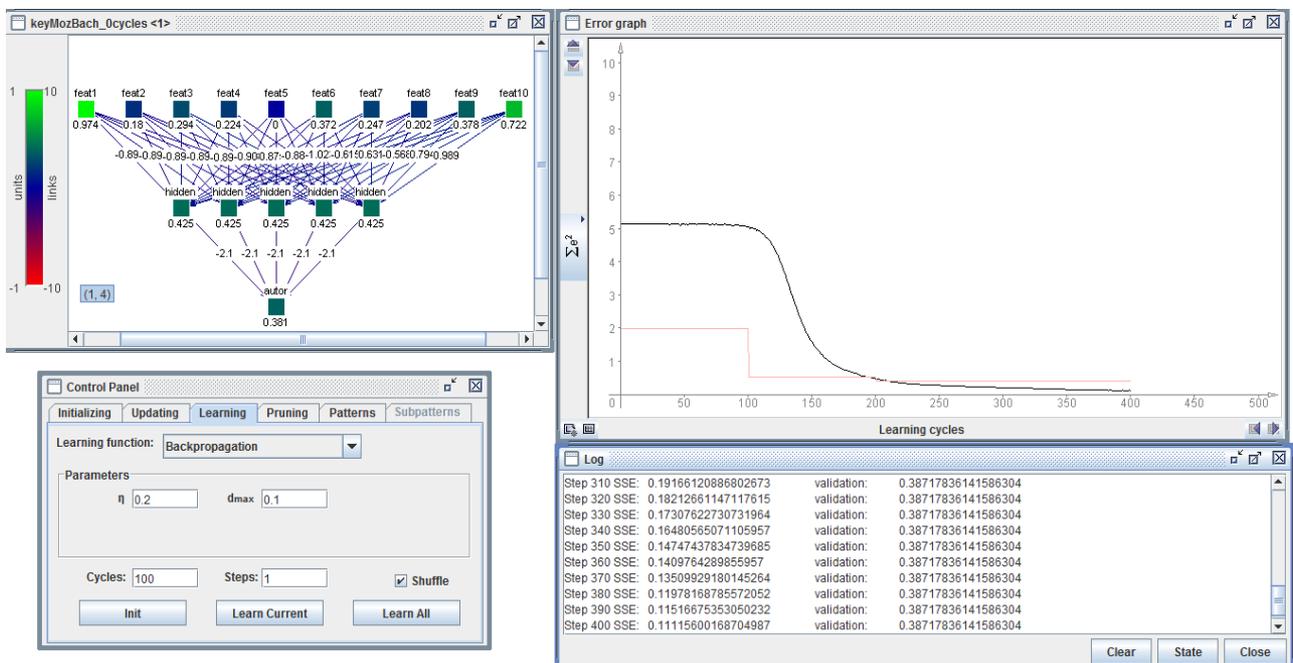


Figura: Bach vs Mozart, primo feature set, rete neuronale dopo 400 cicli di apprendimento

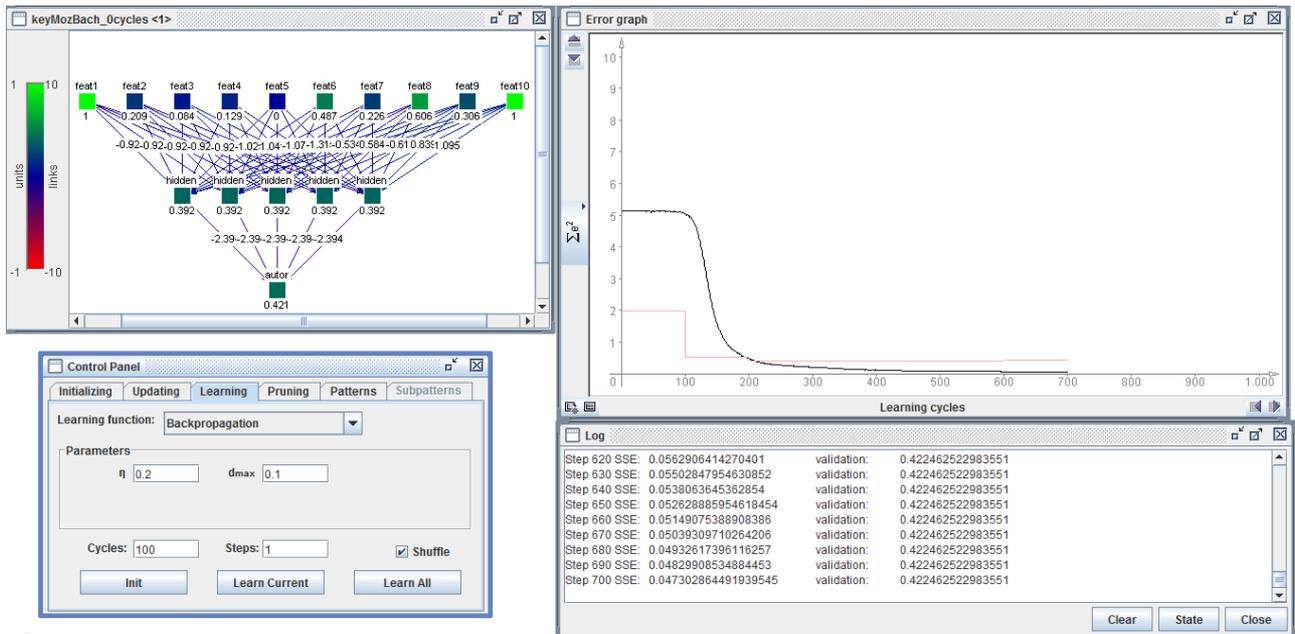


Figura: Bach vs Mozart, primo feature set,
rete neuronale dopo 700 cicli di apprendimento

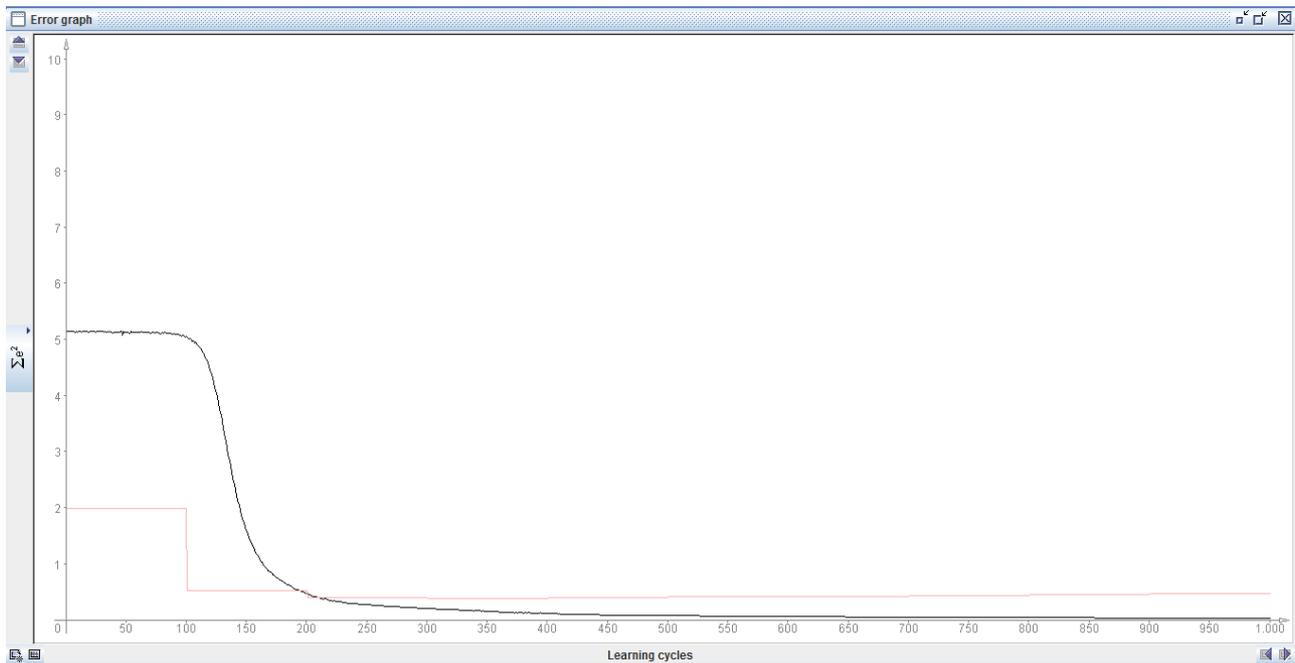


Figura: Bach vs Mozart, primo feature set,
Error Graph della rete neuronale dopo 1000 cicli di apprendimento

La linea nera che rappresenta la somma degli errori al quadrato del training set raggiunge in 200 cicli la linea rossa per poi assestarsi sotto di essa su valori sempre più bassi che tendono a 0.

La linea del validation set continua a scendere, anche se di poco, fino a 400 cicli e raggiunto il minimo comincia gradualmente a risalire, mantenendo poi una lieve pendenza crescente.

La teoria sugli algoritmi di apprendimento suggerisce che **la rete con le migliori capacità di generalizzazione, quindi la più efficiente nel riconoscimento dei due compositori, è quella in cui la linea del validation set raggiunge il minimo**, in questo caso a 400 cicli con $\sum e^2 = 0,3871$. La somma degli errori relativi al training set infatti tende sempre più a 0, ma la rete continua ad essere **addestrata con i soli pattern appartenenti al training set** ed è sempre più "abituata" a riconoscere solo quei particolari pattern, **perdendo gradualmente capacità di generalizzazione**; è per questo che la linea del validation set sale sempre più dopo il raggiungimento del minimo.

Si considera poi il secondo set di feature.

	A	B	C	D	E	F	G	H	I	J	K	Prev
1		Average Note To Note Dynam	Average Time Between A	Average Variability	Duration of Melod	Importance of Bass R	Importance of Middle	Melodic Fifth	Melodic Third	Most Common Pitch	Most Common Pitch	
2	Bach	1.025280	0.604426	0.289753	2.378000	0.897258	1.852273	0.236160	1.086472	0.544829	0.895920	
3	Bach	1.862574	0.462057	0.759637	1.701000	0.704857	2.063804	0.648933	1.266874	0.625027	1.003105	
4	Bach	0.000000	0.608473	0.372037	2.411000	0.732448	1.949393	0.324669	0.892807	0.572788	0.972519	
5	Bach	1.074833	0.649307	0.596964	1.522000	0.657401	2.118250	1.003376	1.125719	0.611048	0.977121	
6	Bach	2.804000	0.760039	0.592286	1.928000	0.728033	1.858895	0.351195	0.533425	0.649675	0.453595	
7	Bach	2.247743	0.603690	0.437041	1.922000	0.705225	2.065379	0.308700	0.461689	0.572420	0.873183	
8	Bach	0.000000	0.696396	0.938009	2.284000	0.974145	1.644053	0.766404	0.614727	0.598908	0.877379	
9	Bach	1.232764	0.602954	0.777772	1.467000	0.629074	2.210955	0.280956	0.795723	0.597068	0.956309	
10	Bach	0.909182	0.940668	0.381859	1.879000	0.657033	2.361284	0.206928	0.461321	0.510985	0.909182	
11	Bach	0.000000	0.550348	0.213018	2.586000	0.695660	2.216842	0.159816	0.401356	0.613255	1.127749	
12	Bach	0.801977	0.771811	1.196635	1.697000	0.546301	2.358107	0.382322	1.288682	0.601483	0.941257	
13	Bach	0.792044	0.591182	0.725938	1.631000	0.777697	1.937253	1.019345	0.387009	0.658504	1.018127	
14	Bach	0.925217	0.642685	0.625655	2.025000	0.654458	2.070793	0.457298	0.429315	0.626131	1.017045	
15	Bach	1.005047	0.555866	0.513868	1.897000	0.849066	2.298511	0.765998	0.653722	0.548508	0.956008	
16	Mozart	2.404092	1.191194	2.228982	1.690000	0.400253	2.035845	0.848552	0.819268	0.710375	1.178094	
17	Mozart	3.399000	1.444663	2.177478	1.850000	0.592654	2.077783	0.705232	0.654090	0.602954	1.028007	
18	Mozart	0.000000	0.483026	1.128155	1.920000	0.576835	1.866620	0.691293	0.777329	0.630913	0.958309	
19	Mozart	2.849000	0.832879	0.763718	1.644000	0.452124	1.915916	0.706721	0.763350	0.768868	1.213551	
20	Mozart	2.936000	0.781744	0.863781	1.578000	0.372662	2.044306	0.920009	0.729137	0.661079	0.952625	
21	Mozart	2.570742	1.060229	2.359211	1.971000	0.476036	1.587032	0.742179	0.652250	0.665494	1.092562	
22	Mozart	3.187675	1.660976	2.785583	1.835000	0.431155	2.130022	0.546755	0.586032	0.654090	1.081058	
23	Mozart	3.044476	1.021105	0.995120	1.498000	0.630913	2.014876	0.579641	0.691981	0.683152	0.967512	
24	Mozart	9.753000	1.137222	1.022052	1.558000	0.606633	1.995378	0.666526	0.733919	0.662919	1.090126	
25	Mozart	3.248347	0.412393	0.411657	1.626000	0.482876	2.017083	0.606437	0.544829	0.735023	1.249957	
26	Mozart	4.083000	1.811438	1.122032	2.285000	0.850169	2.156877	0.434156	0.451020	0.790941	1.340631	
27	Mozart	3.386330	1.504627	2.195872	1.497000	1.346586	2.111996	0.808628	0.585296	0.676898	1.026924	
28	Mozart	7.078000	1.033005	0.529746	1.258000	0.886175	2.614887	0.432626	1.141162	0.807495	0.516503	
29	Mozart	6.251000	0.733919	1.236964	1.805000	1.351999	1.898626	1.026383	0.917491	0.727298	1.295565	

Figura: Bach vs Mozart, secondo set di feature

Come si evince dalla tabella le seconde 10 feature scelte sono:

- "Average Note To Note Dynamics Change", media in unità di velocity MIDI

- del cambiamento di volume tra una nota e la successiva nello stesso canale;
- **"Average Time Between Attacks For Each Voice"**, media dei tempi medi in secondi tra gli eventi "Note On" in ogni singolo canale che contiene almeno una nota;
 - **"Average Variability of Time Between Attacks For Each Voice"**, media delle deviazioni standard del tempo in secondi tra gli eventi "Note On" in ogni singolo canale che contiene almeno una nota;
 - **"Duration of Melodic Arcs"**, numero medio di note che separano i picchi ed i cali melodici in ogni canale;
 - **"Importance of Bass Register"**, quantità di note presenti nell'intervallo dei toni MIDI corrispondenti ai numeri dal 0 al 54;
 - **"Importance of Middle Register"**, quantità di note presenti nell'intervallo dei toni MIDI corrispondenti ai numeri dal 55 al 72;
 - **"Melodic Fifths"**, quantità di intervalli melodici che sono quinte perfette;
 - **"Melodic Thirds"**, quantità di intervalli melodici che sono terze maggiori o minori;
 - **"Most Common Pitch Class Prevalence"**, quantità di eventi di Note On corrispondenti alla *classe della nota* più frequente, dove per *classe di nota* si intende l'insieme di tutte le note con lo stesso nome a prescindere dall'ottava di appartenenza;
 - **"Most Common Pitch Prevalence"**, quantità di eventi di Note On corrispondenti alla nota più frequente;

La rete neuronale costruita è analoga a quella della prova precedente.

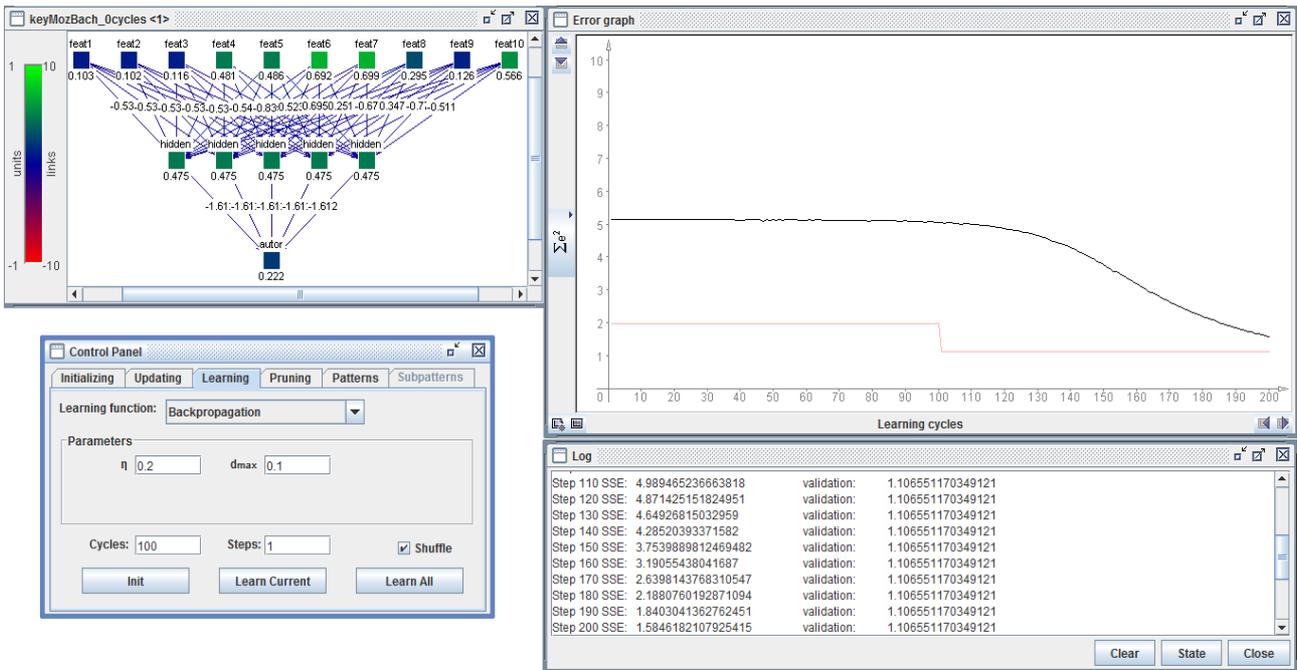


Figura: Bach vs Mozart, secondo feature set, rete neuronale dopo 200 cicli di apprendimento

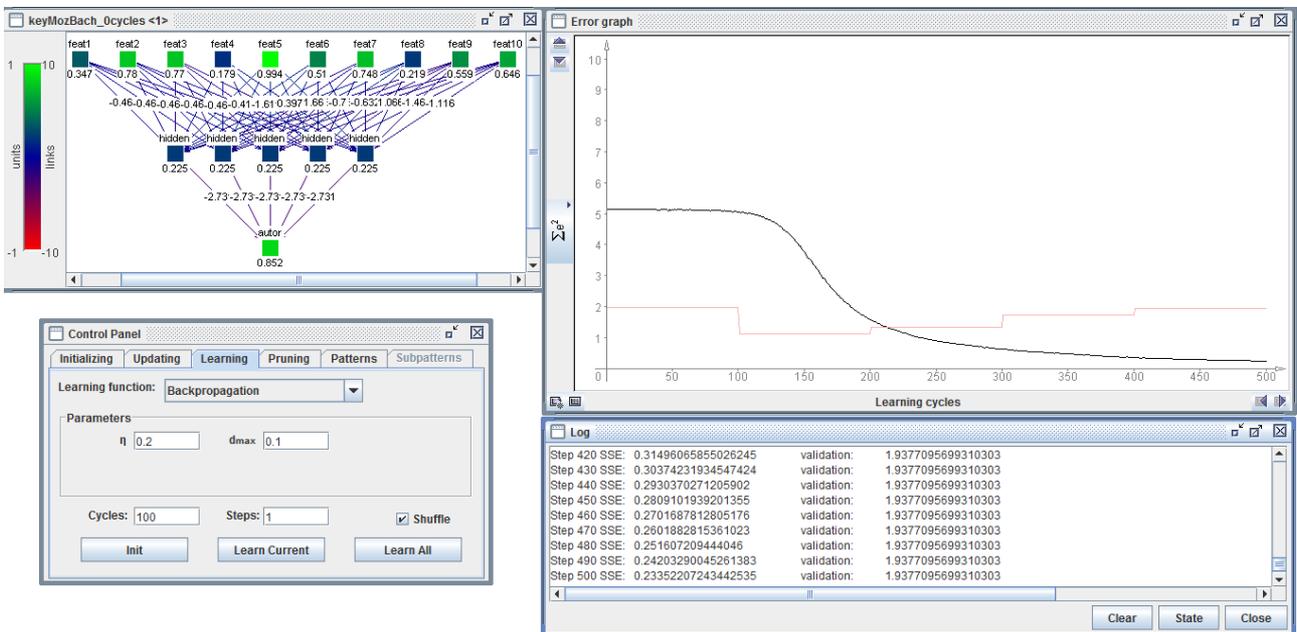


Figura: Bach vs Mozart, secondo feature set, rete neuronale dopo 500 cicli di apprendimento

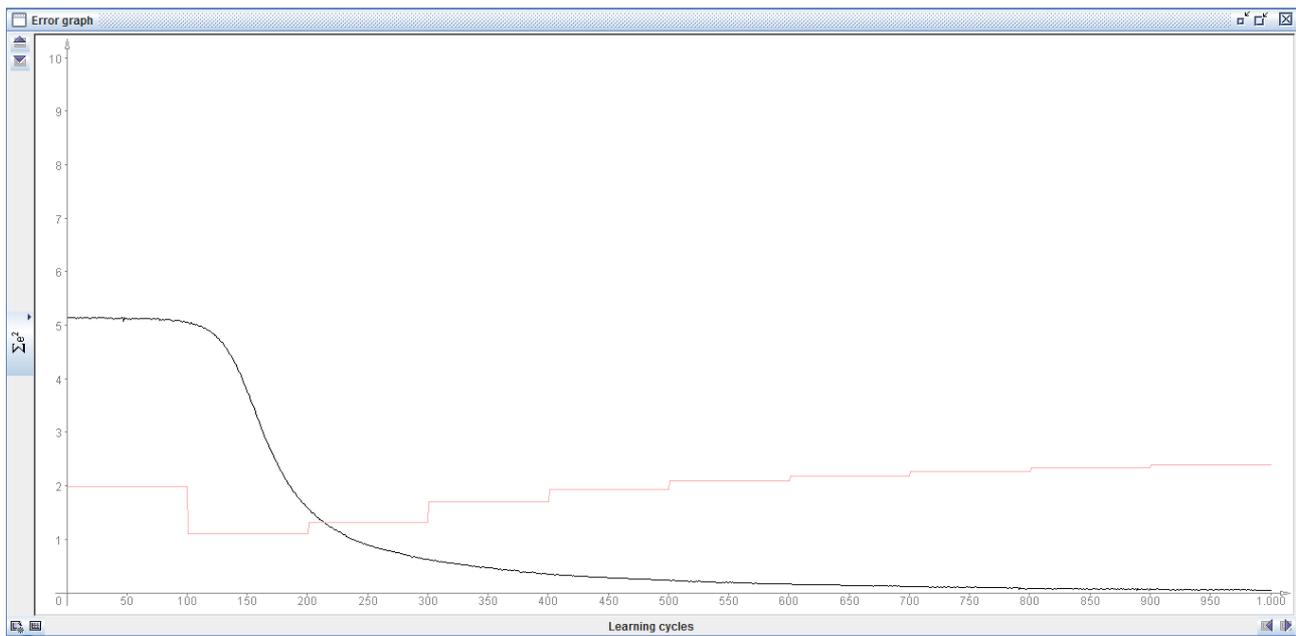


Figura: Bach vs Mozart, secondo feature set,
Error Graph della rete neuronale dopo 1000 cicli di apprendimento

La rete raggiunge il massimo livello di apprendimento dopo 200 cicli. La linea rossa del validation set assume infatti dopo 200 cicli il valore minimo $\sum e^2 = 1,1065$.

Si nota che comunque il valore minimo del secondo set di feature è significativamente più alto del minimo del primo set (che si ricorda essere $\sum e^2 = 0,3871$). Questo porta a concludere che con grande probabilità la classifica delle 20 feature "migliori" è stata costruita secondo criteri sensati, poichè le prime 10 hanno fornito un risultato più efficiente delle seconde 10.

Infine si procede con la costruzione di un set che raccoglie la lista di tutte e 20 le feature selezionate.

La rete neuronale corrispondente presenta questa volta 20 ingressi, quindi 20 unit nello strato di input, 10 unit nello strato hidden e sempre 1 unit nello strato di output.

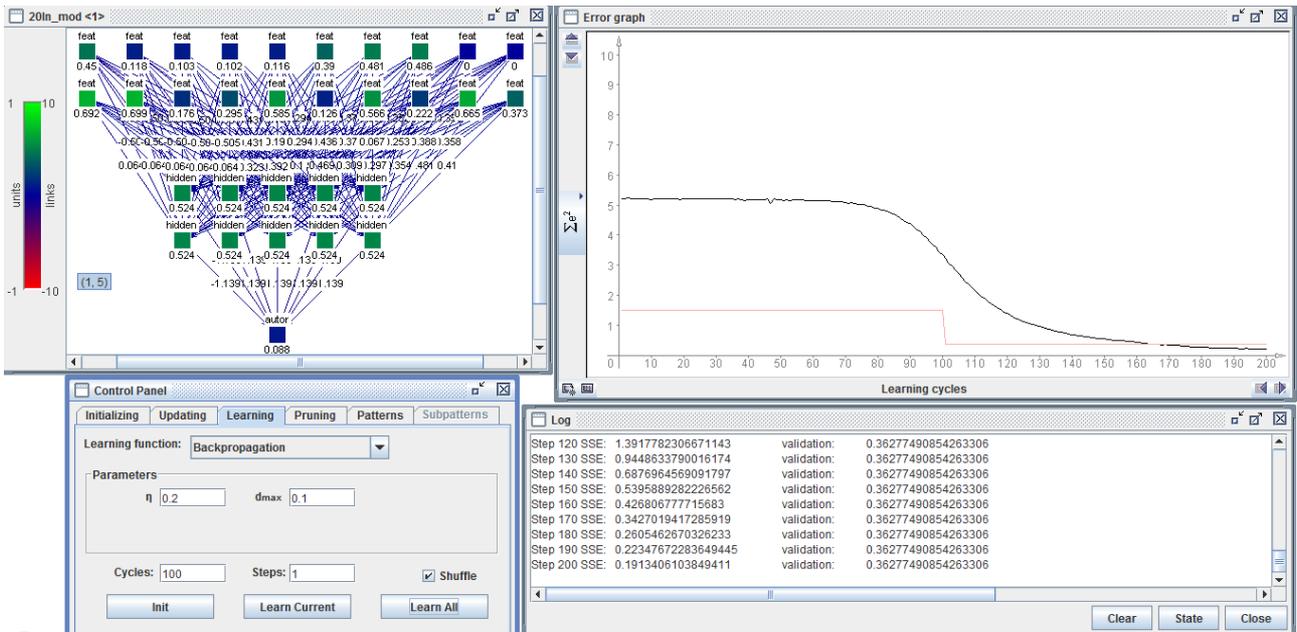


Figura: Bach vs Mozart, terzo feature set, rete neurale dopo 200 cicli di apprendimento

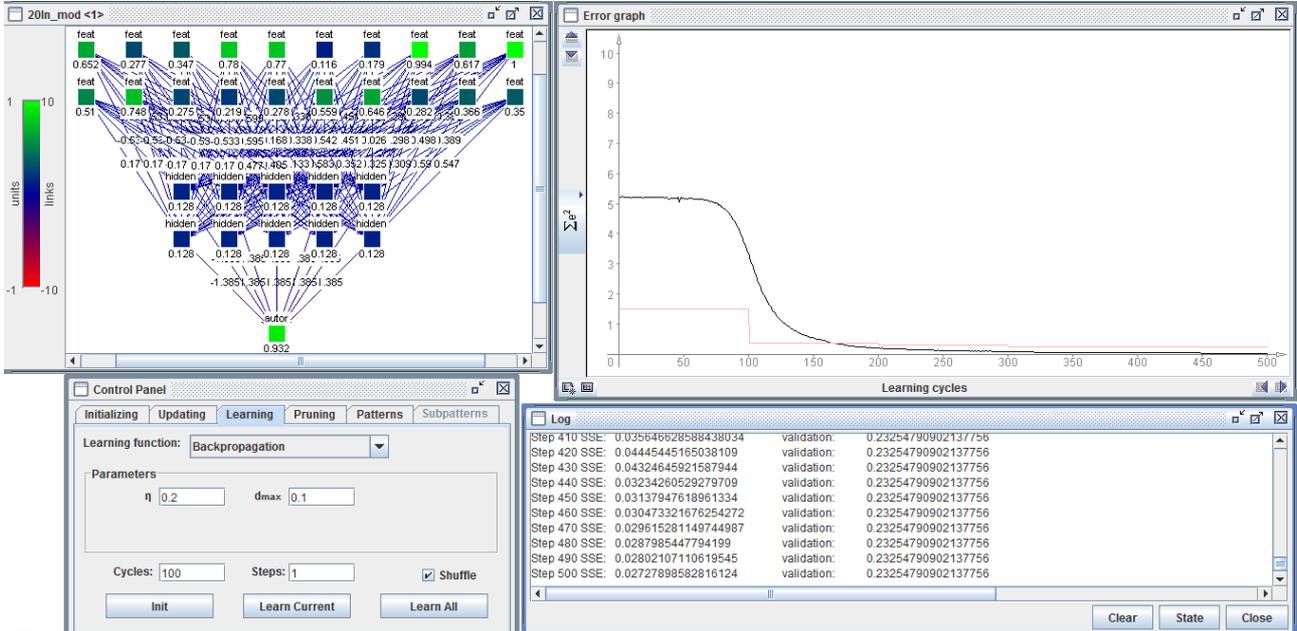


Figura: Bach vs Mozart, secondo feature set, rete neurale dopo 500 cicli di apprendimento

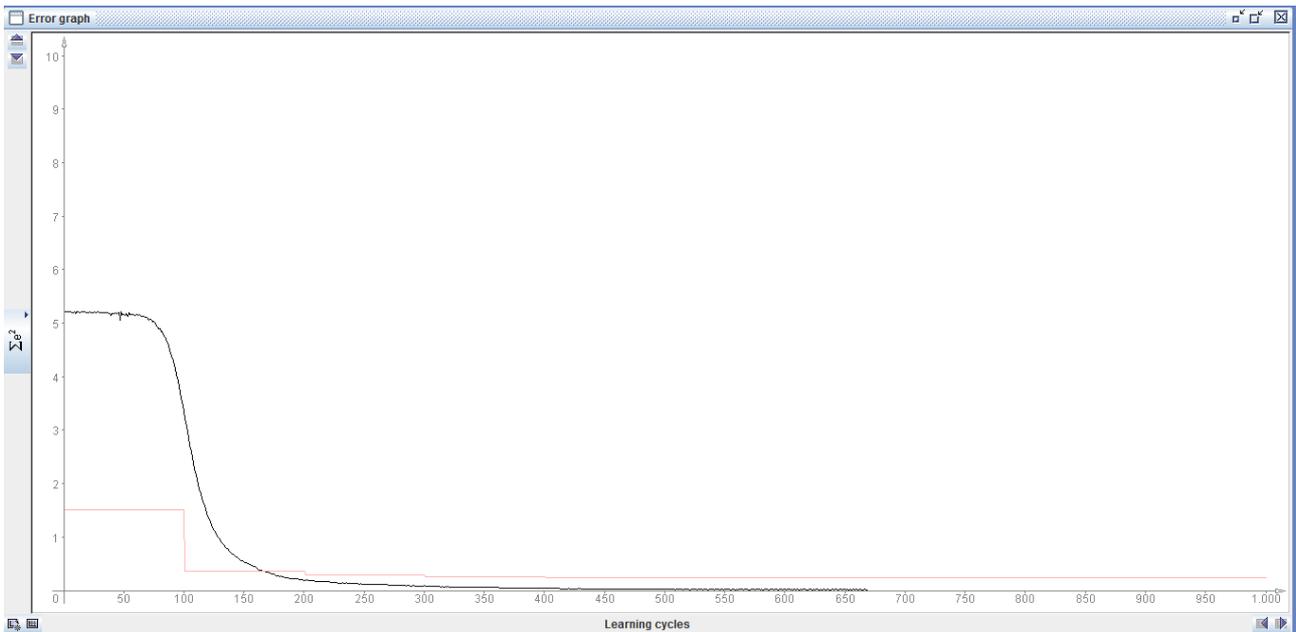


Figura: Bach vs Mozart, terzo feature set,
Error Graph della rete neuronale dopo 1000 cicli di apprendimento

Si nota che utilizzando tutte e 20 le feature il risultato è ancora migliore dell'esperimento con il primo set. Il minimo della linea del validation set infatti raggiunge un valore pari a $\sum e^2 = 0,2325$ dopo. L'apprendimento però dopo circa 670 cicli si ferma poichè la linea nera del training set raggiunge lo 0 e la rete non può essere addestrata ulteriormente.

Complessivamente i risultati di questi tre esperimenti sono piuttosto soddisfacenti, anche se per controparte bisogna dire che è molto facile individuare caratteristiche distintive tra due compositori dallo stile così definito come Bach e Mozart.

6.2

Chet Baker e Luis Armstrong

Come secondo insieme di esperimenti si mettono a confronto due grandi trombettisti Chet Baker e Luis Armstrong. Si sono trovati in questo interessantissimo sito http://pubcs.free.fr/jg/jazz_trumpet_transcriptions_jacques_gilbert_english.html numerosi file MIDI che riproducono accuratamente temi e soli eseguiti da alcuni tra i più importanti trombettisti jazz. Per la grande disponibilità di file si scelgono appunto i due musicisti sopracitati.

Questi esperimenti hanno una valenza diversa rispetto a quelli relativi a Bach e Mozart. Infatti si ha a che fare con dei **brani di improvvisazione**, ad eccezione dei brevi frangenti in cui i musicisti espongono il tema degli standard. L'improvvisazione è definibile come una **composizione istantanea** avulsa da ferree regole esecutive. **Ogni grande jazzista ha uno stile riconoscibile, una propria voce.**

L'obiettivo degli esperimenti di questo paragrafo è di cercare di trovare attraverso la selezione di determinate feature alcune caratteristiche distintive delle voci di Chet Baker e Luis Armstrong.

Si sono scelti 20 tra i più famosi standard suonati da Chet Baker:

- Autumn in New York;
- Stella by Starlight;
- Time on my Hands;
- Well You Needn't;
- Tangerine;
- Minor Yours;
- Star Eyes;

- Autumn Leaves;
- In Your Own Sweet Way;
- But not for Me;
- Happy Little Sunbeam;
- Out of Nowhere;
- Let's Get Lost;
- Look for the Silver Lining;
- Summertime;
- There Will Never Be Another You;
- Blue Room;
- Long Ago and Far Away;
- I Fall in Love Too Easily;
- Pent Up House.

Ecco invece i 20 brani scelti tra i più famosi standard suonati da Luis Armstrong:

- I'm Confessin';
- I'm Not Rough;
- Melancholy Blues;
- Muggles;
- Some of These Days;
- Tight Like This;
- Weary Blues;
- Willie the Weeper;
- You're Driving Me Crazy;
- If I Could Be With You;
- Home;
- Cornet Chop Suey;

- Ain't Misbehavin';
- Heebie Jeebies;
- Gully Low Blues;
- I Double Dare You;
- Hotter than That;
- Hello Dolly;
- Down in Honky Tonky Town;
- Dinah.

A questo punto si scelgono 16 feature giudicate le più significative per cogliere le differenze tra gli stili dei due autori.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	ChetBaker	Amount of Arpeggiation	Average Melodic Direction	Dominant	Duration of Glissando	Melodic Triads	Melodic Triads	Most Common	Note Density	Number of Chords	Number of Phrases	Variability	Variability of Voice Equal	Voice Separation			
2	ChetBaker	1.826154	8.451000	2.100959	6.000000	1.461000	0.000000	0.459114	0.603460	2.718282	3.280966	1.000000	4.000000	2.241857	1.112000	21.993618	2.718282
3	ChetBaker	1.764350	7.204000	2.267977	7.000000	1.750000	0.000000	0.650779	0.798208	3.000000	9.516705	1.000000	5.000000	1.141530	3.399942	13.863237	9.500000
4	ChetBaker	1.729401	8.444000	1.824682	6.000000	1.473000	0.000000	0.561752	0.186763	1.000000	3.501147	0.000000	3.000000	1.339817	3.053031	14.594455	4.756993
5	ChetBaker	1.425533	8.376000	1.723515	2.000000	1.626000	0.000000	0.428947	0.244416	1.000000	4.664572	0.000000	4.000000	0.977824	3.168178	24.089414	2.900407
6	ChetBaker	1.651411	8.059000	1.781272	7.000000	1.553000	0.000000	0.473093	0.211394	2.000000	5.042413	0.000000	3.000000	0.854952	3.060757	16.869657	3.941509
7	ChetBaker	1.367776	8.459000	1.756624	7.000000	1.587000	0.000000	0.421590	0.426577	2.000000	4.879316	0.000000	3.000000	0.921170	2.961062	13.765379	4.213337
8	ChetBaker	1.469310	7.472000	1.784583	5.000000	1.643000	0.000000	0.593390	0.387600	2.000000	4.221492	0.000000	3.000000	1.183468	2.963637	13.966532	4.213337
9	ChetBaker	1.768764	6.515000	1.988756	7.000000	1.732000	0.000000	0.815221	0.840026	3.000000	7.366544	1.000000	5.000000	1.829464	1.150000	18.473443	2.718282
10	ChetBaker	1.495430	7.100000	1.741173	6.000000	1.655000	0.000000	0.523860	0.545537	1.000000	3.599005	0.000000	3.000000	1.302293	3.022130	14.458451	4.077423
11	ChetBaker	1.608369	9.472000	1.834983	5.000000	1.489000	0.000000	0.385906	0.310594	2.000000	4.275857	0.000000	4.000000	1.094809	2.914341	21.632087	3.805595
12	ChetBaker	1.539208	8.257000	1.795620	5.000000	1.485000	0.000000	0.384066	0.350518	1.000000	4.776021	0.000000	3.000000	0.889532	2.972466	15.249561	4.349251
13	ChetBaker	1.514192	8.138000	1.752210	5.000000	1.574000	0.000000	0.451388	0.263768	1.000000	3.199418	0.000000	4.000000	2.036581	2.996010	8.401357	9.333000
14	ChetBaker	1.493223	9.345000	1.736759	7.000000	1.490000	0.000000	0.428580	0.452967	1.000000	3.248347	0.000000	3.000000	1.464160	3.017347	21.822367	4.213337
15	ChetBaker	1.453492	8.803000	1.708432	6.000000	1.528000	0.000000	0.395103	0.388277	1.000000	4.061113	0.000000	3.000000	1.097752	3.004839	16.898840	4.213337
16	ChetBaker	1.922538	6.891000	1.987285	4.000000	1.761000	0.000000	0.608105	1.005812	5.000000	4.368279	0.000000	4.000000	1.429947	1.022000	16.899558	2.718282
17	ChetBaker	1.663551	7.967000	1.746692	6.000000	1.526000	0.000000	0.484865	0.184868	1.000000	4.471574	1.000000	3.000000	0.971202	3.039788	15.382757	3.941509
18	ChetBaker	1.548037	8.417000	1.802241	6.000000	1.544000	0.000000	0.368983	0.600664	5.000000	4.607488	0.000000	4.000000	1.038156	3.067379	19.645023	3.079813
19	ChetBaker	1.968991	6.304000	2.024441	4.000000	1.784000	0.000000	0.738702	0.538228	5.000000	5.175609	0.000000	4.000000	1.516031	3.650468	12.335563	2.990110
20	ChetBaker	1.835351	7.102000	2.087348	6.000000	1.494000	0.000000	0.474932	0.518710	6.000000	3.223882	0.000000	5.000000	2.484658	1.135000	10.025023	7.250000
21	ChetBaker	1.252997	7.729000	1.827825	4.000000	1.592000	0.000000	1.190815	0.322233	2.000000	4.463419	0.000000	3.000000	1.137851	3.013301	9.652619	3.941509
22	LuisArmstrong	1.927688	7.800000	2.310283	5.000000	1.922000	0.000000	0.842076	1.126396	7.000000	4.846697	2.000000	6.000000	1.502420	1.497000	8.032523	8.750000
23	LuisArmstrong	2.026280	5.181000	2.213898	2.000000	1.729000	0.228523	1.053239	0.498169	3.000000	4.504193	2.000000	5.000000	0.723619	1.242000	5.743730	7.750000
24	LuisArmstrong	1.660608	2.734592	1.782376	3.000000	1.535000	0.000000	0.388849	0.751111	7.000000	4.939118	0.000000	5.000000	0.810071	1.405000	4.929245	9.000000
25	LuisArmstrong	1.875449	6.595000	2.056078	3.000000	1.740000	0.177640	0.754153	0.807275	3.000000	6.045459	1.000000	6.000000	0.916388	1.424000	3.522893	8.750000
26	LuisArmstrong	2.081462	7.707000	2.547197	5.000000	2.013000	0.000000	0.976720	0.447709	3.000000	10.875846	1.000000	5.000000	0.872242	1.387000	11.612500	8.500000
27	LuisArmstrong	2.046145	4.970000	2.128918	3.000000	1.732000	0.000000	0.836926	0.749893	2.000000	4.585741	2.000000	5.000000	0.767630	1.256000	7.741667	8.500000
28	LuisArmstrong	1.957854	4.793000	2.202862	3.000000	1.832000	0.000000	0.866724	0.737983	3.000000	7.165391	1.000000	5.000000	0.404667	1.214000	8.892000	9.750000
29	LuisArmstrong	2.091027	7.369000	2.490176	3.000000	2.104000	0.084190	0.931471	1.072938	4.000000	10.718185	3.000000	5.000000	0.813381	1.404000	3.357078	9.750000
30	LuisArmstrong	1.943507	7.787000	2.461113	3.000000	2.054000	0.096505	0.921538	1.237100	4.000000	9.948911	1.000000	5.000000	0.626131	1.445000	3.792003	9.000000
31	LuisArmstrong	2.238914	7.212000	2.258044	7.000000	1.615000	0.000000	0.826625	0.974008	3.000000	4.218773	1.000000	5.000000	1.609105	1.144000	7.866708	9.250000
32	LuisArmstrong	1.962637	7.561000	2.194401	5.000000	1.862000	0.165152	0.985181	0.509673	3.000000	4.868443	1.000000	4.000000	2.051296	1.353000	4.525939	3.079813
33	LuisArmstrong	1.825050	7.470000	2.214266	2.000000	1.755000	0.000000	0.859734	0.786433	3.000000	6.855507	0.000000	4.000000	0.707064	1.369000	6.763085	3.172235
34	LuisArmstrong	1.642950	6.174000	2.043938	6.000000	1.599000	0.071046	0.590447	0.597436	6.000000	8.285323	1.000000	5.000000	0.504731	1.732000	3.985001	9.250000
35	LuisArmstrong	2.151727	7.685000	2.484658	6.000000	2.003000	0.000000	1.055078	0.420854	3.000000	8.799078	3.000000	5.000000	0.714054	1.434000	6.928900	8.000000
36	LuisArmstrong	1.992803	6.583000	1.991699	3.000000	1.854000	0.000000	0.984445	0.800779	3.000000	6.871816	2.000000	5.000000	1.161395	1.297000	4.868443	8.250000
37	LuisArmstrong	1.990964	7.951000	2.478404	5.000000	2.018000	0.000000	0.971937	0.501052	3.000000	10.927493	1.000000	4.000000	0.737230	1.356000	10.788881	8.750000
38	LuisArmstrong	1.824682	6.709000	2.085141	5.000000	1.790000	0.000000	0.872242	0.788869	3.000000	9.872900	0.000000	5.000000	0.603690	1.351000	4.324786	9.250000
39	LuisArmstrong	2.281220	4.979000	2.320584	6.000000	1.976000	0.000000	1.417072	0.587220	3.000000	7.668273	1.000000	6.000000	0.498845	1.275000	5.193000	9.500000
40	LuisArmstrong	2.025544	7.060000	2.456699	3.000000	2.010000	0.000000	0.840237	0.504363	6.000000	12.406238	2.000000	4.000000	0.542254	1.409000	5.200073	8.750000
41	LuisArmstrong	2.167178	7.495000	2.445663	2.000000	2.043000	0.107311	1.087820	1.032879	4.000000	11.832681	3.000000	5.000000	0.588239	1.398000	8.994795	9.500000

Figura: Baker vs Armstrong, tabella contenente il set di feature

Come si evince dalla figura le feature scelte sono le seguenti:

- **Amount of Arpeggiation**, quantità di intervalli orizzontali che sono note ripetute, terze minori, terze maggiori, quinte giuste, settime minori, settime maggiori, ottave, decime minori o decime maggiori;

- **Average Melodic Interval**, media degli intervalli melodici espressa in numero di semitoni;
- **Direction of Motion**, quantità di intervalli melodici ascendenti, piuttosto che discendenti;
- **Dominant Spread**, maggior numero di *classi di note* consecutive separate da intervalli di quinta perfetta;
- **Duration of Melodic Arcs**, numero medio di note che separano i picchi ed i cali melodici in ogni canale;
- **Glissando Prevalence**, numero di eventi Note On che hanno almeno un MIDI Pitch Bend associato ad essi diviso il numero totale di eventi Note On;
- **Melodic Thirds**, quantità di intervalli melodici che sono terze maggiori o minori;
- **Melodic Tritones**, quantità di intervalli melodici che sono tritoni;
- **Most Common Melodic Interval**, l'intervallo melodico con la maggiore frequenza;
- **Note Density**, media di note per secondo;
- **Number of Common Pitches**, numero di note che rappresentano da sole almeno il 9% di tutte le note;
- **Number of Pitched Instruments**, numero totale di tracce MIDI usate per riprodurre almeno una nota;
- **Variability of Note Duration**, deviazione standard della durata delle note in secondi;
- **Variability of Number of Independent Voices**, deviazione standard del numero di canali differenti in cui le note sono eseguite simultaneamente (con resti esclusi dal calcolo);
- **Voice Equality - Note Duration**, deviazione standard della durata totale delle note in secondi in ogni canale che contiene almeno una nota;
- **Voice Separation**, separazione media in semitoni tra le medie dei toni di canali

consecutivi che contengono almeno una nota.

In un genere come il jazz è sicuramente più complicato cogliere dei pattern caratteristici o comunque non è affatto semplice codificarli. Tuttavia ascoltando Chet Baker e Luis Armstrong si notano facilmente alcune differenze di stile significative.

Baker è riflessivo e "pesa" ogni nota, spesso indugiando a lungo su alcune semplici, ma efficacissime frasi. Mentre suona pare che mediti e cerchi risposte a complesse domande intime, interiori.

Armstrong invece nei suoi soli è **irruento**, a tratti viscerale, tuttavia molto preciso.

Usa più note nelle frasi e adotta molti abbellimenti ed effetti quali il glissando. Ne risulta uno stile piuttosto diretto, privo di esitazioni e dal ritmo serrato.

Queste differenze sono riscontrabili nella tabella di feature presentata.

Le feature "Amount of Arpeggiation" e "Note Density" presentano valori più alti per Armstrong a riprova del suo stile ricco di note, di scale veloci ed arpeggi.

"Variability of Note Duration" e "Voice Equality - Note Duration" presentano, soprattutto la seconda, valori nettamente più alti per Baker. Ciò è riconducibile al suo maggior utilizzo di note lunghe e meditate.

E' interessante notare anche il maggior impiego di terze maggiori e minori e di tritoni da parte di Armstrong; così come valori più alti per la feature "Most Common Melodic Interval". Questo è dovuto in parte ad una maggiore variabilità nelle frasi di Chet Baker, ma soprattutto alla frequenza di note più alta di Armstrong che nello stesso periodo di tempo esplora mediamente nelle sue frasi molti più intervalli.

Inoltre si riscontra l'uso del glissando da parte di Armstrong a differenza di Baker con la feature "Glissando Prevalence".

Altre feature riguardanti il glissando e la durata delle note estratte da JSymbolic sono chiaramente significative, ma si è deciso di scartarle perchè giudicate ridondanti.

A questo punto si costruisce la rete neuronale in JavaNNS. Essa presenta 16 ingressi, 8 unit nello strato hidden e 1 unit nello strato di output.

Si costruiscono i file ".pat" con 14 pattern per autore (quindi 28 in totale) nel file di training set e 6 pattern per autore (quindi 12 in totale) nel validation set.

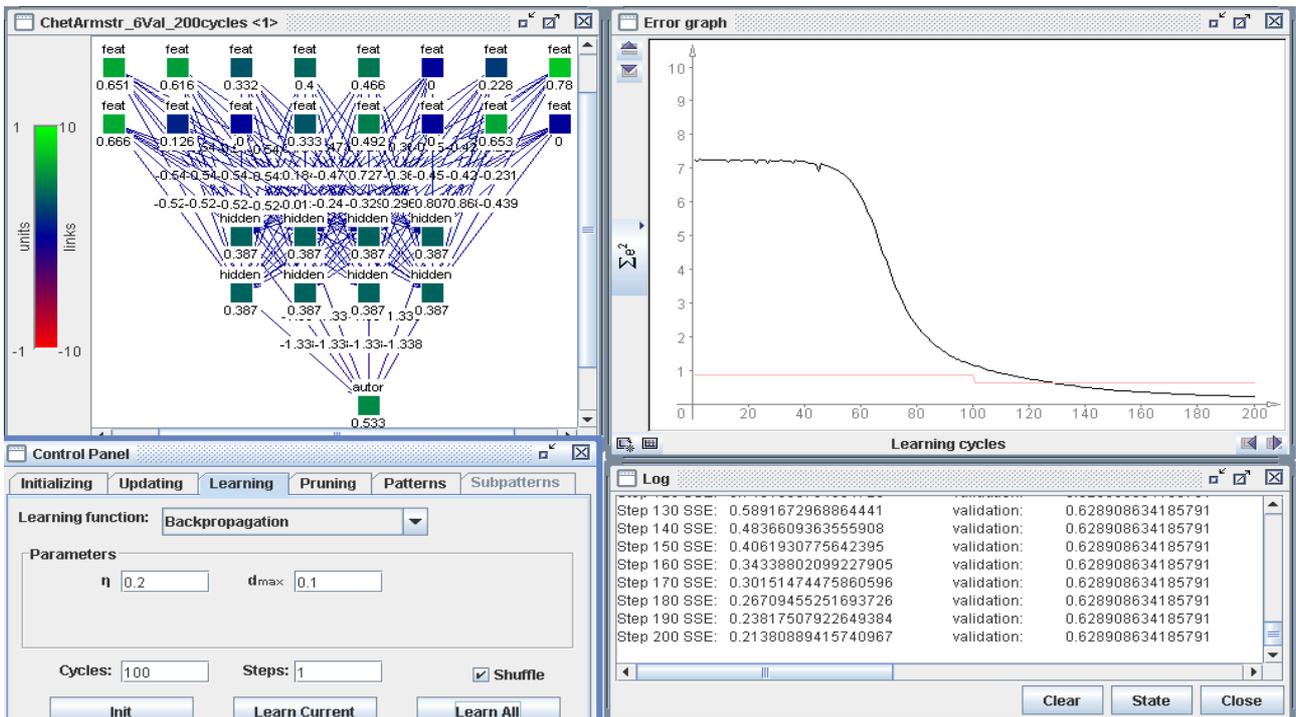


Figura: Baker vs Armstrong, 28 pat training 12 pat validation ,
rete neuronale dopo 200 cicli di apprendimento

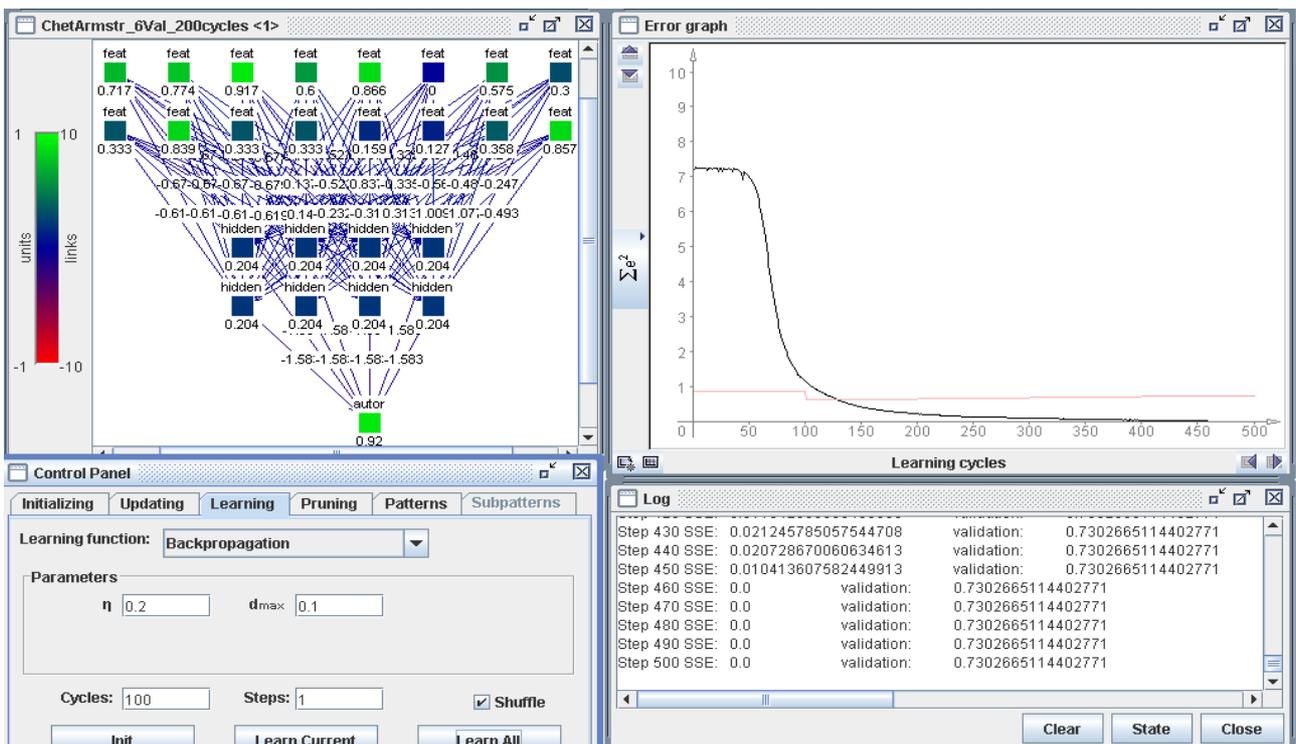


Figura: Baker vs Armstrong, 28 pat training 12 pat validation ,
rete neuronale dopo 500 cicli di apprendimento

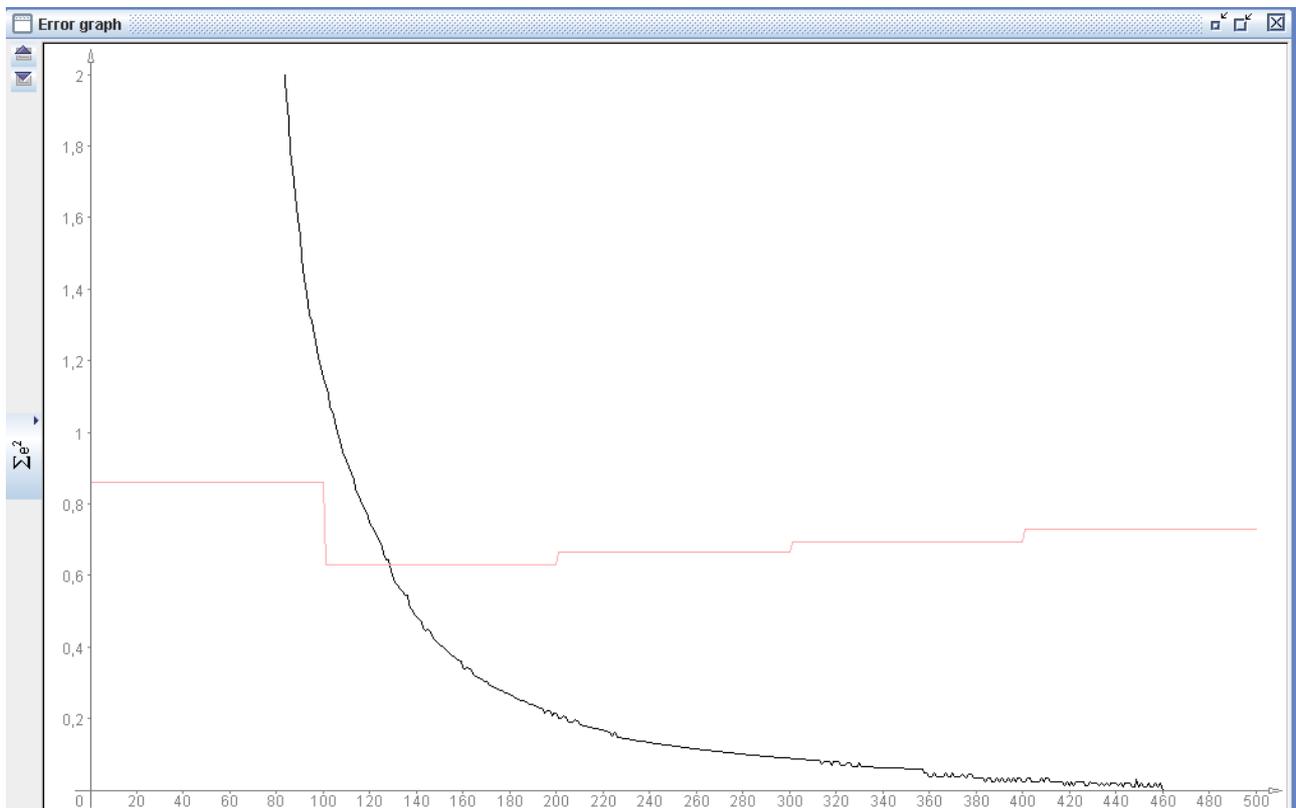


Figura: Baker vs Armstrong, 28 pat training 12 pat validation,
Error Graph della rete neuronale dopo 500 cicli di apprendimento

La rete ha raggiunto un certo livello di apprendimento e il suo picco, corrispondente al minimo della funzione errore al quadrato del validation set, è $\sum e^2 = 0,6289$.

Si sperimenta in seguito una nuova struttura di rete neuronale, riducendo a 4 il numero di unit nello strato hidden, per verificare se sussistono differenze nel livello di apprendimento.

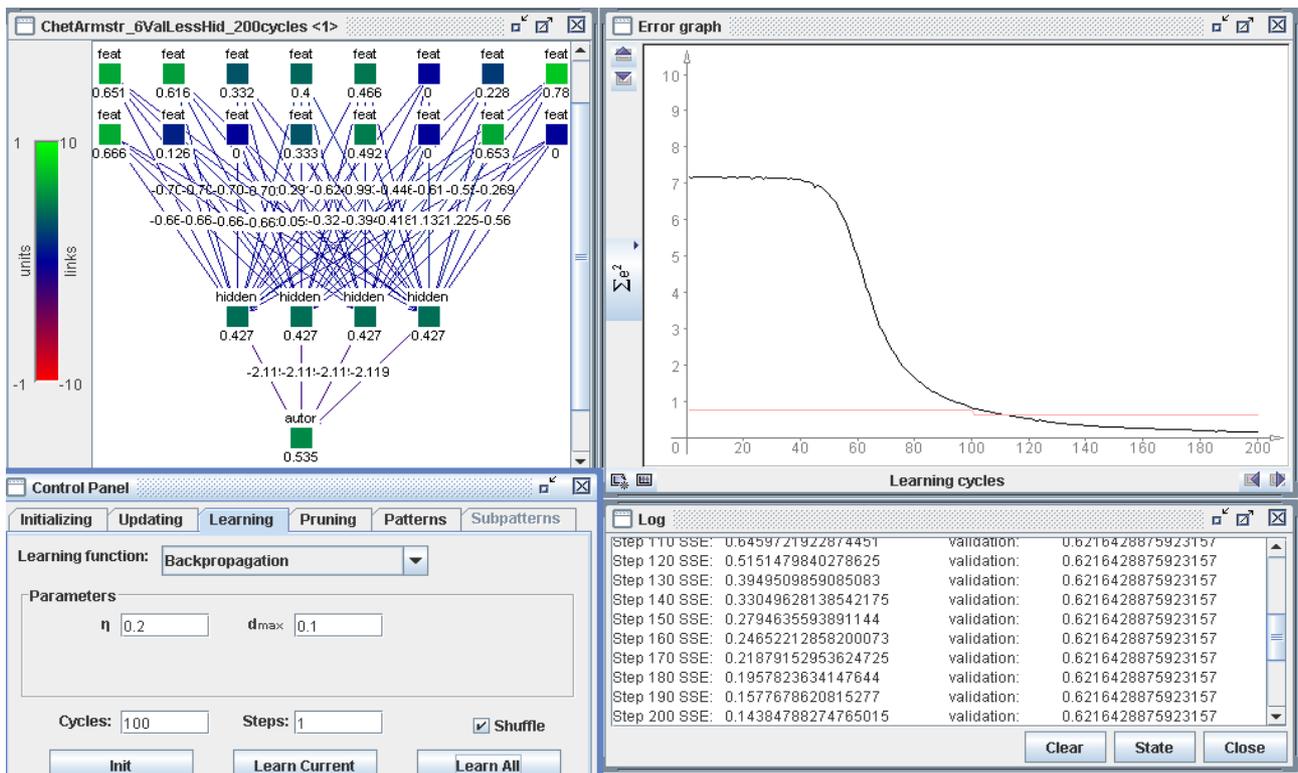


Figura: Baker vs Armstrong, 28 pat training 12 pat validation ,
rete neuronale con 4 unit hidden dopo 200 cicli di apprendimento

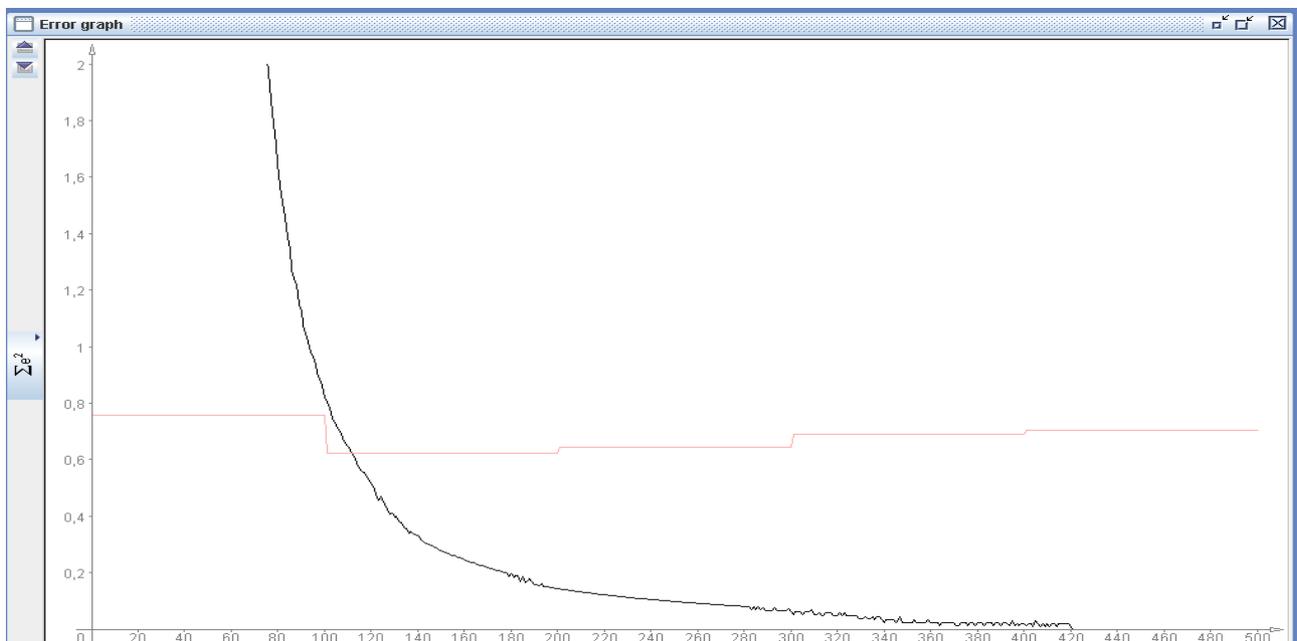


Figura: Baker vs Armstrong, 28 pat training 12 pat validation ,
Error Graph della rete neuronale con 4 unit hidden dopo 500 cicli di apprendimento

Come si nota sussiste una differenza nel livello di apprendimento se si usano 4 unit nello strato hidden invece di 8, ma essa è veramente minima, praticamente trascurabile; $\sum e^2 = 0,6216$ invece di $\sum e^2 = 0,6289$ dopo 200 cicli.

Si prova poi a modificare la distribuzione dei pattern tra il training ed il validation set, impiegando 12 pattern per autore (quindi 24 in totale) nel file di training set e 8 pattern per autore (quindi 16 in totale) nel validation set. Si utilizza per questo esperimento la rete con 8 unit hidden.

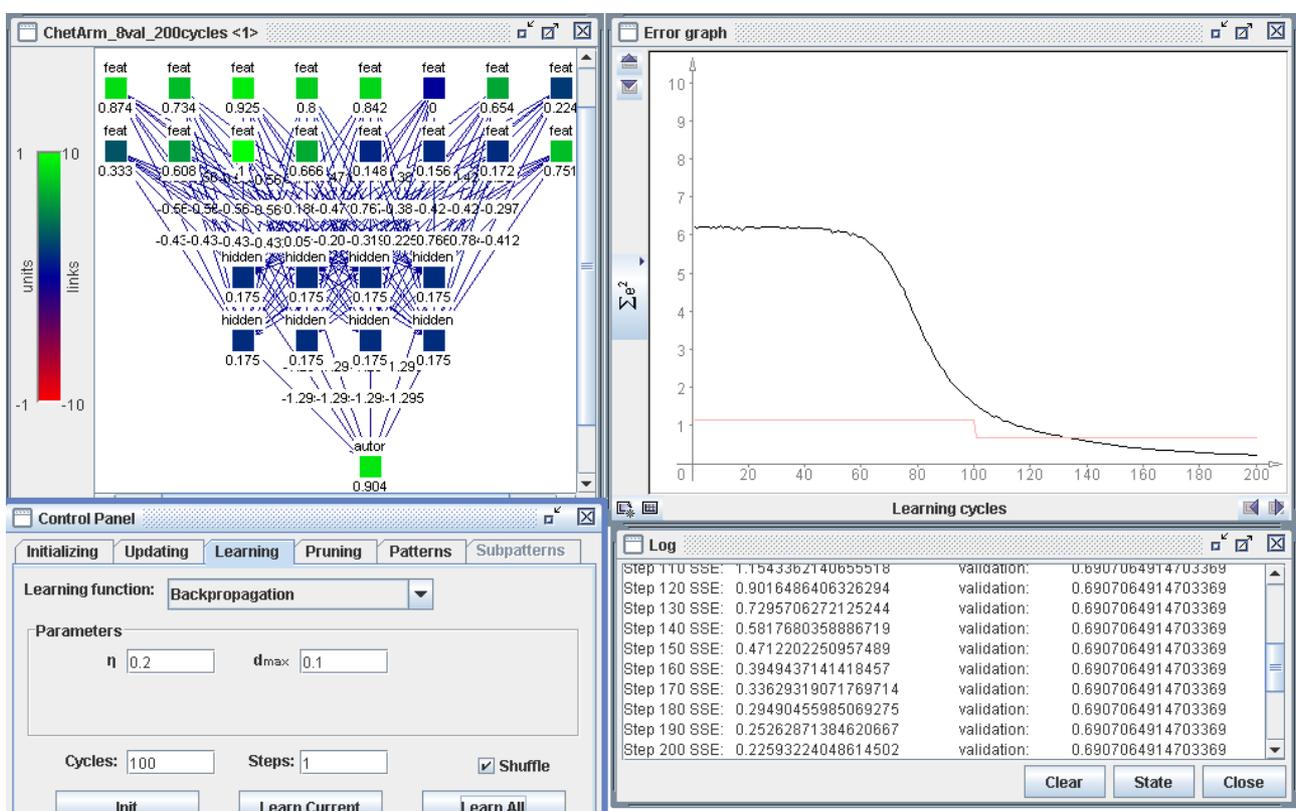


Figura: Baker vs Armstrong, 24 pat training 16 pat validation ,
rete neuronale dopo 200 cicli di apprendimento

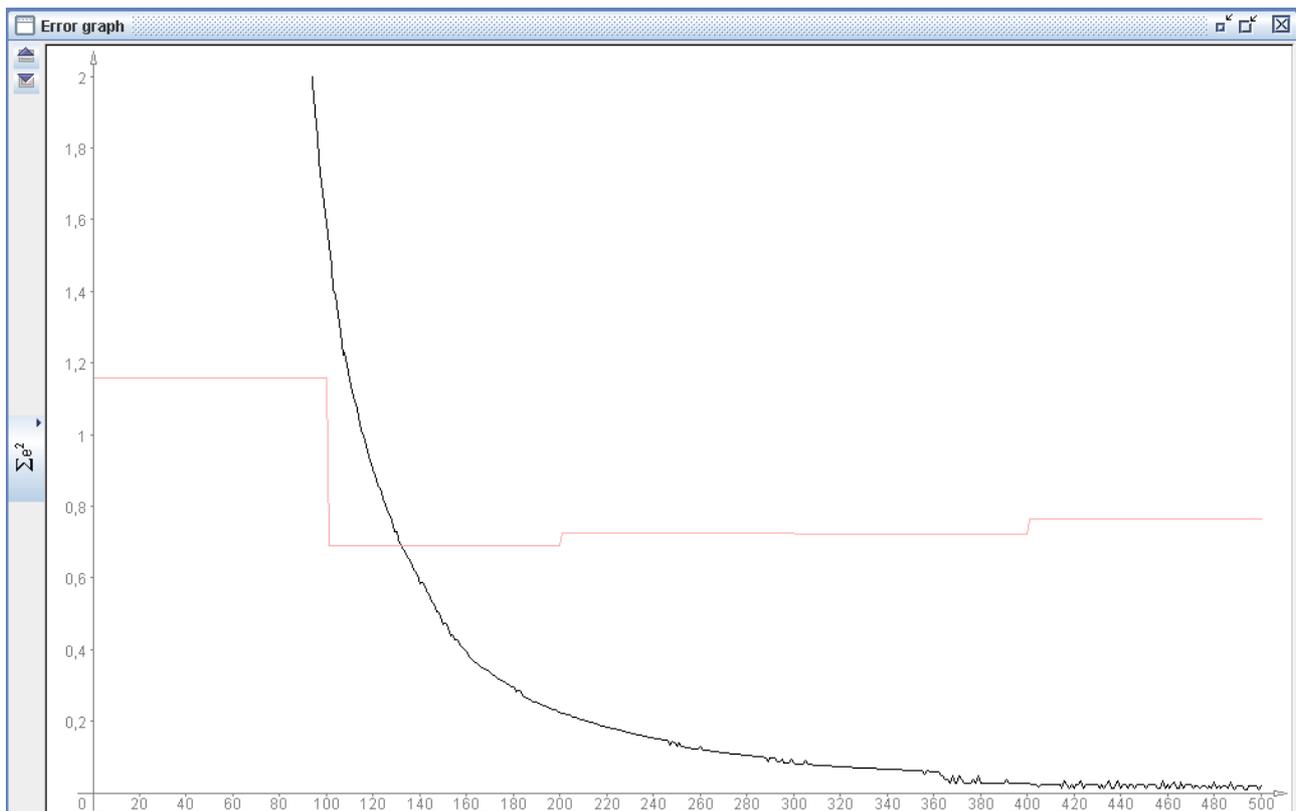


Figura: Baker vs Armstrong, 24 pat training 16 pat validation ,
Error Graph della rete neurale dopo 500 cicli di apprendimento

Si nota che il minimo ottenuto con questa nuova distribuzione è $\sum e^2 = 0,6907$ raggiunto dopo 200 cicli di apprendimento.

Si procede nell'analisi degli stili di Chet Baker e Luis Armstrong con un approccio diverso.

Si decide di confrontare gli stessi brani dei due autori con l'aggiunta di altri 4, mettendoli a confronto con brani casuali di altri trombettisti generici.

Si costruiscono cioè per ogni esperimento due insiemi di brani. Il primo contenente tutti pezzi o di Chet Baker o di Luis Armstrong e il secondo contenente standard suonati da diversi trombettisti jazz scelti casualmente.

I brani scelti casualmente sono eseguiti dai seguenti musicisti: **B. Berigan, B. Beiderbecke, B. Clayton, B. Hackett, C. Shavers, R. Eldridge e W. Vache.**

Ci si accerta di costruire l'insieme in modo che presenti lo stesso numero di esempi per ognuno di essi.

Il primo ad essere confrontato con questa lista di trombettisti è Chet Baker.

Si sono scelte le stesse 16 feature degli esperimenti precedenti che confrontavano Baker ed Armstrong direttamente.

Si sono aggiunti 4 brani di Chet Baker ai 20 precedentemente elencati e scelti 24 brani scelti casualmente con la modalità descritta sopra.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	ChetBa	Amount of A	Average Melo	Direction of P	Dominant*	Duration of P	Glissando*	Melodic Thir*	Melodic Trit*	Most Comm*	Note Density	Number of O	Number of P	Variability of P	Variability*	Voice Equal*	Voice Separation
2	ChetBa	1,826154	8,451000	2,100959	6,000000	1,461000	0,000000	0,459114	0,603460	2,718282	3,280966	1,000000	4,000000	2,241857	1,112000	21,993618	2,718282
3	ChetBa	1,764350	7,204000	2,267977	7,000000	1,750000	0,000000	0,650779	0,798208	3,000000	9,516705	1,000000	5,000000	1,141530	3,399942	13,863237	9,500000
4	ChetBa	1,729401	8,444000	1,824682	6,000000	1,473000	0,000000	0,561752	0,186763	1,000000	3,501147	0,000000	3,000000	1,339817	3,053031	14,594455	4,756993
5	ChetBa	1,425533	8,376000	1,723515	2,000000	1,626000	0,000000	0,428947	0,244416	1,000000	4,664572	0,000000	4,000000	0,977824	3,168178	24,089414	2,900407
6	ChetBa	1,651411	8,059000	1,781272	7,000000	1,553000	0,000000	0,473093	0,211394	2,000000	5,042413	0,000000	3,000000	0,854952	3,060757	16,869657	3,941509
7	ChetBa	1,367776	8,459000	1,756624	7,000000	1,587000	0,000000	0,421590	0,426577	2,000000	4,879316	0,000000	3,000000	0,921170	2,961062	13,765379	4,213337
8	ChetBa	1,469310	7,472000	1,784583	5,000000	1,643000	0,000000	0,593390	0,387600	2,000000	4,221492	0,000000	3,000000	1,183468	2,963637	13,966532	4,213337
9	ChetBa	1,768764	6,515000	1,988756	7,000000	1,732000	0,000000	0,815221	0,840026	3,000000	7,366544	1,000000	5,000000	1,829464	1,150000	18,473443	2,718282
10	ChetBa	1,495430	7,100000	1,741173	6,000000	1,655000	0,000000	0,523860	0,545537	1,000000	3,599005	0,000000	3,000000	1,302293	3,022130	14,458441	4,077423
11	ChetBa	1,608369	9,472000	1,834983	5,000000	1,489000	0,000000	0,385906	0,310594	2,000000	4,275857	0,000000	4,000000	1,094809	2,914341	21,632087	3,805595
12	ChetBa	1,539208	8,257000	1,795620	5,000000	1,485000	0,000000	0,384066	0,350518	1,000000	4,776021	0,000000	3,000000	0,889532	2,972466	15,249561	4,349251
13	ChetBa	1,514192	8,138000	1,752210	5,000000	1,574000	0,000000	0,451388	0,263768	1,000000	3,199418	0,000000	4,000000	2,036581	2,996010	8,401357	9,333000
14	ChetBa	1,493223	9,345000	1,736759	7,000000	1,490000	0,000000	0,428580	0,452967	1,000000	3,248347	0,000000	3,000000	1,464160	3,017347	21,822367	4,213337
15	ChetBa	1,453492	8,803000	1,708432	6,000000	1,528000	0,000000	0,395103	0,388277	1,000000	4,061113	0,000000	3,000000	1,097752	3,004839	16,896840	4,213337
16	ChetBa	1,922538	6,891000	1,987285	4,000000	1,761000	0,000000	0,608105	1,005812	5,000000	4,368279	0,000000	4,000000	1,429947	1,022000	16,899558	2,718282
17	ChetBa	1,968891	6,304000	2,024441	4,000000	1,784000	0,000000	0,738702	0,538228	5,000000	5,175609	0,000000	4,000000	1,516031	3,650468	12,335563	2,901110
18	ChetBa	2,148048	6,605000	2,042835	4,000000	1,697000	0,000000	0,967155	0,588438	4,000000	5,257157	0,000000	4,000000	1,287578	3,594182	3,737638	3,079813
19	ChetBa	1,663551	7,967000	1,746692	6,000000	1,526000	0,000000	0,484865	0,184868	1,000000	4,471574	1,000000	3,000000	0,971202	3,039788	15,382757	3,941509
20	ChetBa	1,766925	4,895000	1,865149	4,000000	1,838000	0,000000	0,867460	0,354578	1,000000	7,591000	0,000000	4,000000	0,885118	3,407667	3,772975	2,718282
21	ChetBa	1,543622	2,205000	1,945346	3,000000	2,000000	0,000000	1,280956	0,000000	2,000000	2,568000	4,000000	1,000000	0,728033	0,000000	0,000000	0,000000
22	ChetBa	1,783480	2,758000	2,452652	2,000000	1,737000	0,000000	1,672012	0,000000	2,000000	4,250000	4,000000	1,000000	0,425637	0,000000	0,000000	0,000000
23	ChetBa	2,043938	2,980000	1,576731	5,000000	1,542000	0,000000	1,597700	0,000000	4,000000	2,273000	3,000000	1,000000	0,834351	0,000000	0,000000	0,000000
24	ChetBa	1,548037	8,417000	1,802241	6,000000	1,544000	0,000000	0,368983	0,500064	5,000000	4,607488	0,000000	4,000000	1,038156	3,067379	19,645023	3,079813
25	ChetBa	1,252997	7,729000	1,827625	4,000000	1,592000	0,000000	1,190815	0,322233	2,000000	4,463419	0,000000	3,000000	1,137851	3,013301	9,652619	3,941509
26	other	1,896786	6,279000	2,076312	5,000000	1,725000	0,087970	0,780002	1,137222	5,000000	3,965973	1,000000	5,000000	2,224935	1,031000	15,162576	7,500000
27	other	1,965580	7,691000	2,441616	2,000000	2,008000	0,000000	0,975984	1,248333	3,000000	8,149409	2,000000	5,000000	0,911237	1,452000	7,249658	9,000000
28	other	1,778697	7,464000	2,101327	3,000000	1,821000	0,000000	0,805288	1,084171	3,000000	6,822887	0,000000	4,000000	1,045881	1,026000	3,454936	2,807985
29	other	1,571213	9,078000	1,813278	6,000000	1,965000	0,000000	1,013508	0,680330	4,000000	8,570743	2,000000	5,000000	0,909766	1,121000	16,399394	8,500000
30	other	1,747427	3,185000	1,986549	5,000000	1,711000	0,000000	0,721044	0,316549	1,000000	4,569432	0,000000	8,000000	3,348071	4,053000	4,602051	2,786000
31	other	1,804449	6,453000	2,036581	7,000000	1,568000	0,000000	0,553659	0,420118	6,000000	3,278248	0,000000	5,000000	3,185100	1,075000	19,683079	7,000000
32	other	0,896154	3,019000	1,976248	6,000000	2,512000	0,000000	0,586032	0,466405	2,000000	4,735247	0,000000	6,000000	1,351957	3,059000	9,356000	4,400000
33	other	1,373662	7,935000	1,695924	6,000000	1,473000	0,000000	0,373766	0,506560	2,000000	4,800486	0,000000	3,000000	0,983342	3,017715	10,188120	4,349251
34	other	2,133701	4,472000	1,887589	3,000000	2,075000	0,000000	1,460481	0,777501	3,000000	4,175281	2,000000	5,000000	2,939357	2,806552	12,738733	5,333000
35	other	1,955279	7,407000	2,417704	6,000000	2,097000	0,000000	0,967891	1,064277	3,000000	9,236722	1,000000	5,000000	0,650411	1,566000	3,324459	8,000000
36	other	1,726458	7,294000	1,696660	5,000000	1,592000	0,000000	0,643053	0,379615	2,000000	4,808641	0,000000	4,000000	0,967155	3,147576	11,680457	2,990110
37	other	1,800402	6,752000	2,108685	6,000000	1,729000	0,000000	0,898729	0,906882	3,000000	5,102215	1,000000	5,000000	2,292625	1,344000	5,287058	8,000000
38	other	1,537000	3,959000	1,962269	2,000000	1,978000	0,000000	0,776226	0,477408	1,000000	7,059378	2,000000	7,000000	1,121297	4,323000	5,902000	2,214000
39	other	1,694085	9,642000	2,120825	5,000000	1,564000	0,000000	0,853112	0,982534	4,000000	3,819186	0,000000	5,000000	3,072897	1,171000	20,101694	8,250000
40	other	1,539575	7,097000	1,729769	6,000000	1,615000	0,000000	0,600011	0,290430	1,000000	3,267375	0,000000	3,000000	1,488440	2,980927	5,599661	4,621079
41	other	1,670908	6,975000	1,822843	3,000000	1,486000	0,000000	0,609944	0,390307	1,000000	9,390000	1,000000	3,000000	2,124504	2,932735	14,847255	4,485165
42	other	1,901201	7,570000	2,412921	5,000000	2,039000	0,000000	0,979295	0,448445	4,000000	11,386883	2,000000	5,000000	0,575633	1,553000	3,335332	9,500000
43	other	1,997953	7,780000	2,475461	6,000000	2,093000	0,000000	0,962740	1,171327	4,000000	7,785159	1,000000	5,000000	0,765925	1,461000	7,950974	8,250000
44	other	1,466000	8,421000	1,733448	3,000000	1,492000	0,000000	0,564695	0,369719	1,000000	4,892907	0,000000	3,000000	0,912341	3,202758	3,098841	4,485165
45	other	0,949497	3,224000	1,950497	5,000000	2,515000	0,000000	0,678002	0,458887	2,000000	4,786894	0,000000	6,000000	1,339081	3,061000	9,443000	4,400000
46	other	1,419647	7,284000	1,736023	4,000000	1,556000	0,000000	0,580514	0,317361	1,000000	3,164080	0,000000	3,000000	1,413393	2,949657	5,626843	4,621079
47	other	1,406771	9,285000	2,018187	4,000000	1,712000	0,000000	0,693821	0,758960	2,000000	4,890189	0,000000	5,000000	2,273495	1,163000	8,500067	8,000000
48	other	1,987285	3,316000	1,802977	2,000000	1,819000	0,284483	1,289785	0,311136	3,000000	3,017000	5,000000	1,000000	0,668437	0,000000	0,000000	0,000000
49	other	1,709536	6,523000	1,920331	7,000000	1,872000	0,000000	0,809703	0,376341	3,000000	4,039367	0,000000	5,000000	2,223660	1,106000	25,029939	2,786239
50																	

Figura: Baker vs Other, tabella contenente il set di feature

Si impiegano 16 pattern per autore (quindi 32 in totale) nel file di training set e 8 pattern per autore (quindi 16 in totale) nel validation set.

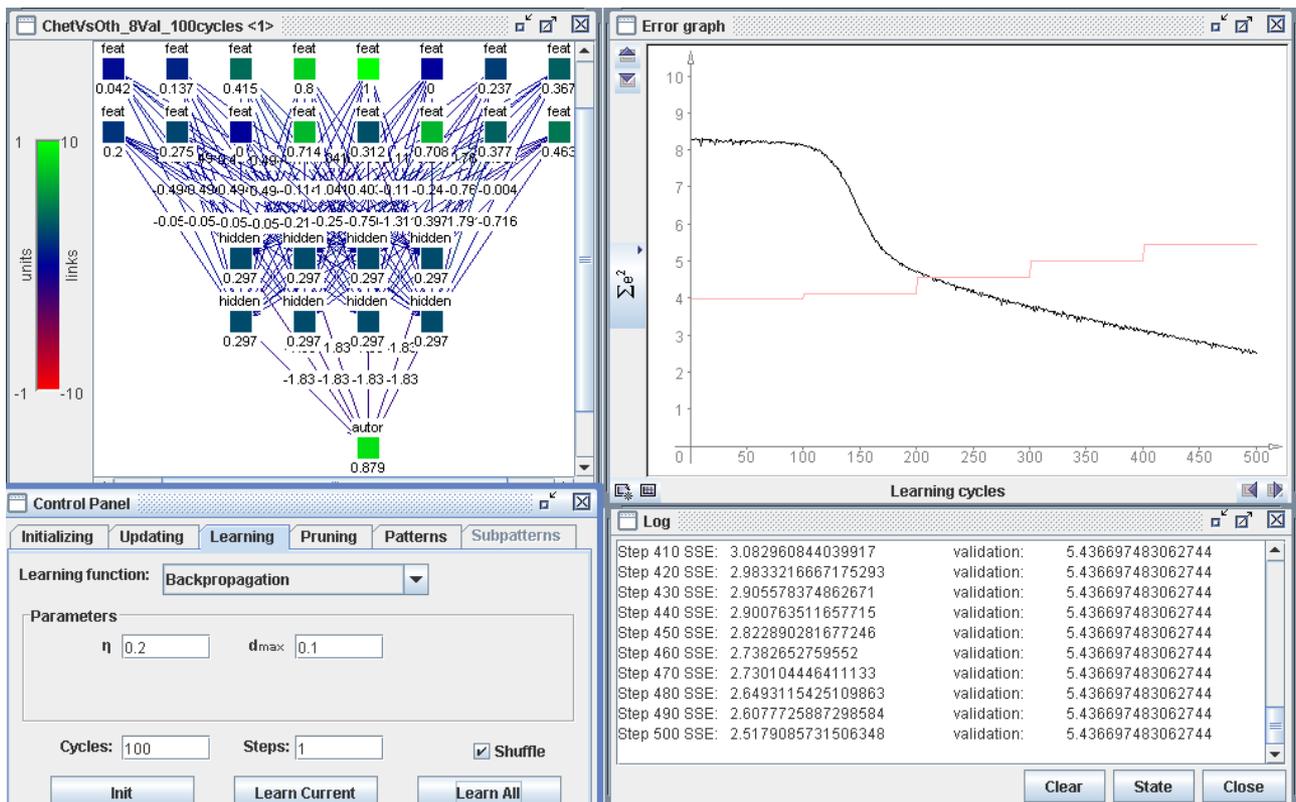


Figura: Baker vs Other, 32 pat training 16 pat validation, $\eta=0.2$, rete neuronale dopo 500 cicli di apprendimento

Si nota che per la prima volta la linea rossa relativa al validation set non scende mai rispetto al valore iniziale. Se ne deduce che la rete neuronale non apprende affatto.

Si prova allora a modificare il parametro di apprendimento η che, come detto nel primo capitolo sulle reti neurali a proposito dell'algoritmo di Backpropagation, indica la distanza di cui ci si deve muovere nella direzione definita dal gradiente nel processo di minimizzazione della funzione errore per scegliere il nuovo punto in cui calcolare iterativamente il gradiente.

Si decide di abbassare il valore di η , da $0,2$ a $0,1$, e ripetere l'esperimento.

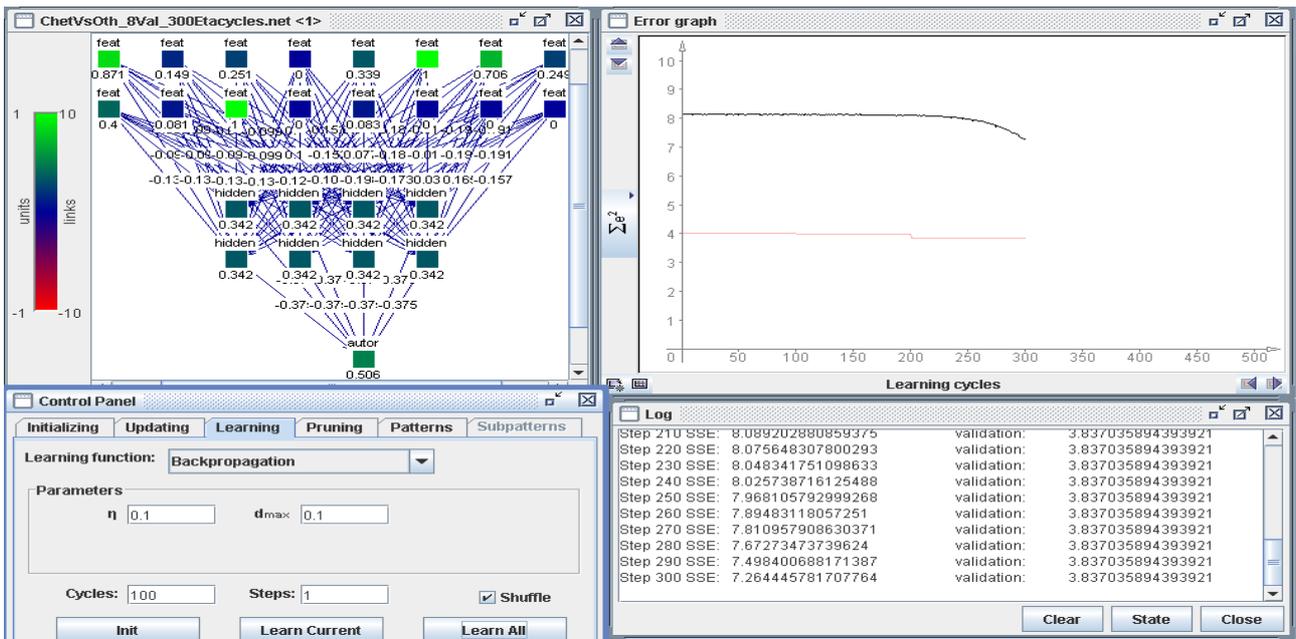


Figura: Baker vs Other, 32 pat training 16 pat validation, $\eta=0.1$, rete neuronale dopo 300 cicli di apprendimento

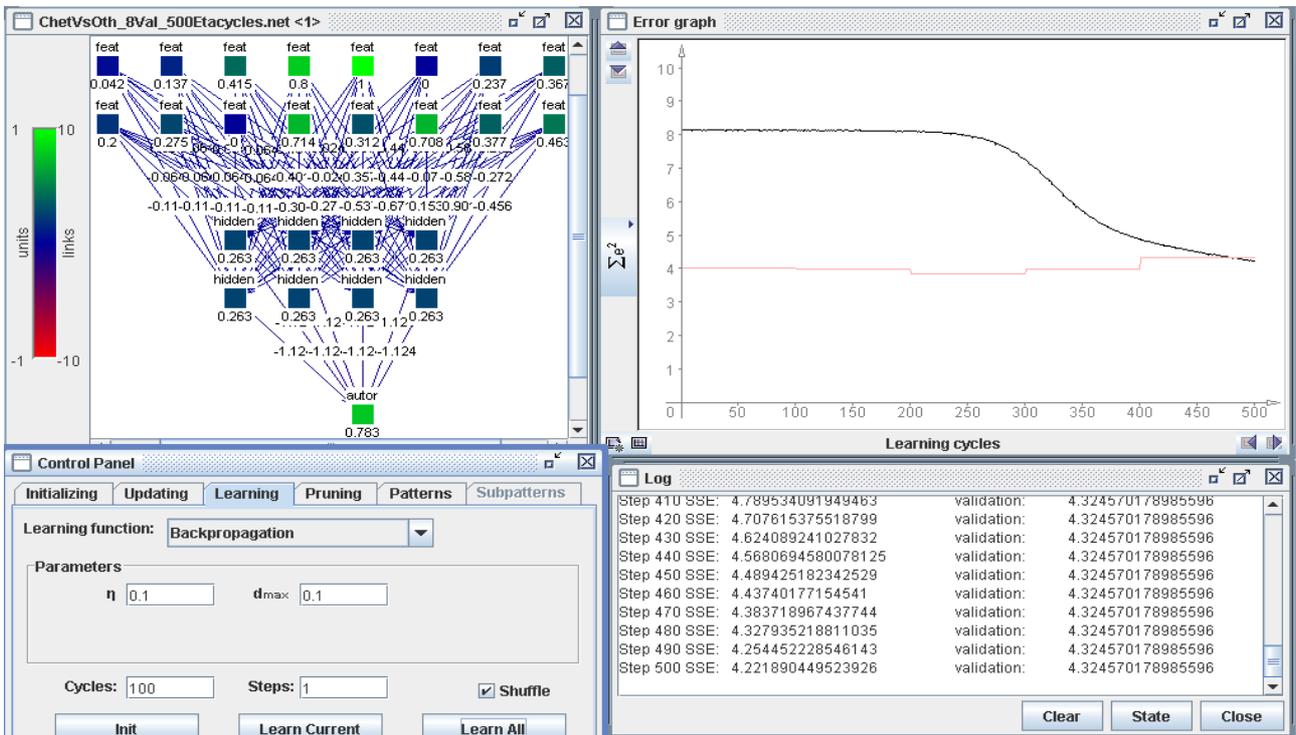


Figura: Baker vs Other, 32 pat training 16 pat validation, $\eta=0.1$, rete neuronale dopo 500 cicli di apprendimento

Si nota che questo cambiamento delle impostazioni dell' algoritmo porta ad un miglioramento delle prestazioni. Si riesce infatti a raggiungere un picco di apprendimento dopo 300 cicli con $\sum e^2 = 3,8370$.

Si procede allora con una ulteriore diminuzione del parametro η , settandolo a $\eta=0,05$ per sperare di assistere di nuovo ad un miglioramento.

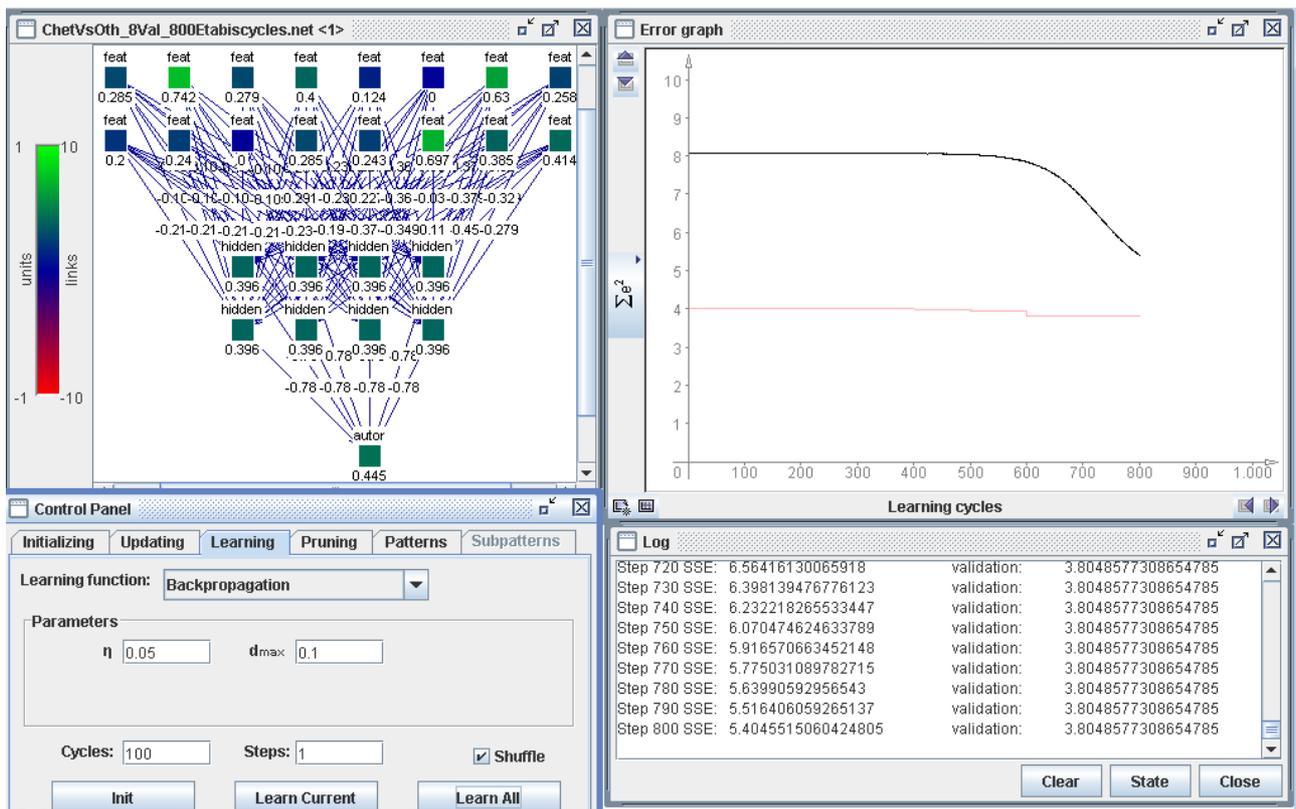


Figura: Baker vs Other, 32 pat training 16 pat validation, $\eta=0,05$,
rete neuronale dopo 800 cicli di apprendimento

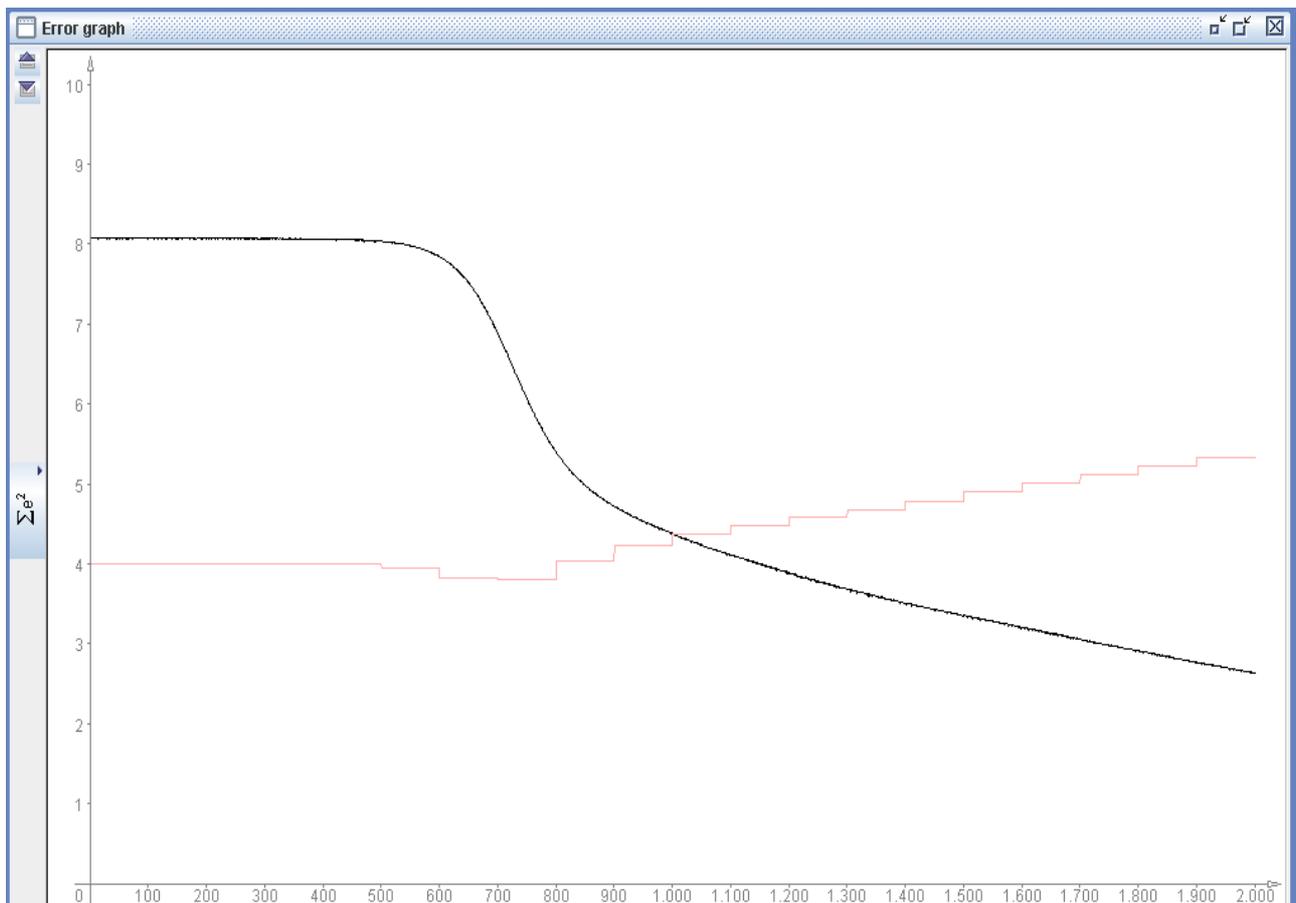


Figura: Baker vs Other, 32 pat training 16 pat validation, $\eta=0.05$,
Error Graph della rete neurale dopo 2000 cicli di apprendimento

Effettivamente il minimo della linea rossa relativa al validation set raggiunge un **valore migliore** del caso precedente, infatti $\sum e^2 = 3,8048$ invece di $\sum e^2 = 3,8370$ dopo 800 cicli di apprendimento.

Bisogna però notare che **per controparte è necessario aspettare molti più cicli di apprendimento per ottenere risposte dalla rete**. Questo è ragionevole dato che, come detto, η diminuisce la distanza di spostamento nel processo di minimizzazione della funzione errore.

Procediamo allora ancora con la diminuzione del parametro η , portandolo a $\eta=0,005$.

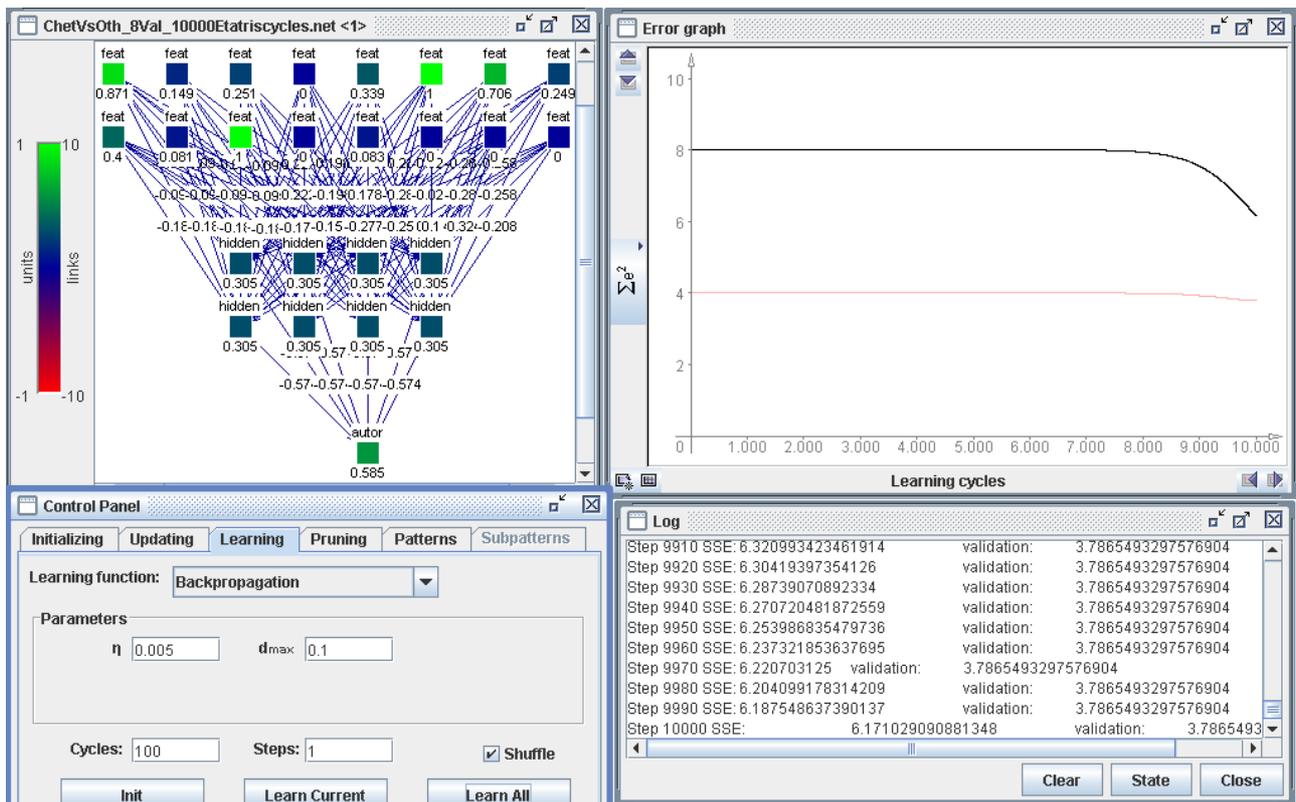


Figura: Baker vs Other, 32 pat training 16 pat validation, $\eta=0.005$,
rete neuronale dopo 10000 cicli di apprendimento

Si nota, coerentemente con le valutazioni fatte in precedenza, che con le nuove impostazioni il processo di apprendimento è ancora più lento .

E' necessario addirittura aspettare 10000 cicli di apprendimento perchè la funzione errore cada in un minimo che è lievemente migliore del precedente $\sum e^2 = 3,7865$ invece di $\sum e^2 = 3,8048$.

Si decide dunque di fermarsi qui con la diminuzione di η .

Si passa poi ad un nuovo esperimento in cui si confronta Armstrong con la lista di trombettisti descritta in precedenza. **Le feature utilizzate sono sempre le stesse 16** impiegate negli esperimenti del confronto diretto tra Baker e Armstrong e negli

esperimenti precedenti di Baker comparato con altri trombettisti.

Si aggiungono 4 brani di Armstrong alla lista di inizio paragrafo, e i 24 brani misti sono gli stessi degli esperimenti con Baker illustrati in precedenza.

Si impiegano 20 pattern per autore (quindi 40 in totale) nel file di training set e 4 pattern per autore (quindi 8 in totale) nel validation set.

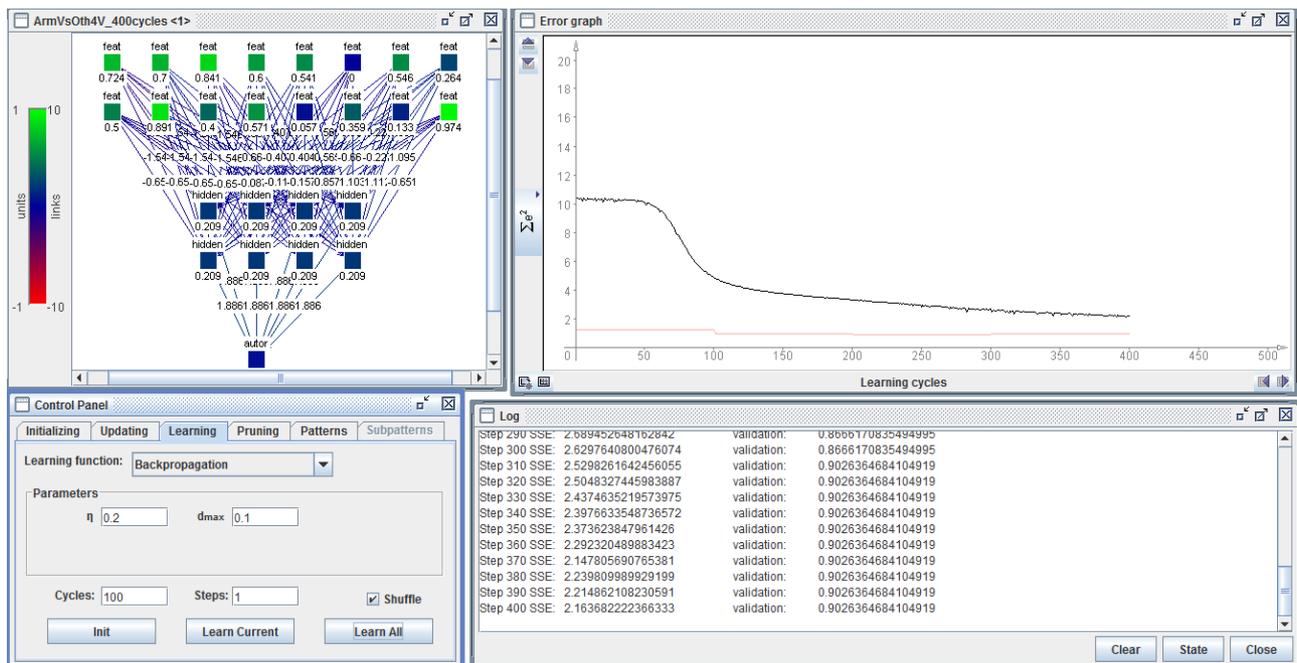


Figura: Armstrong vs Other, 40 pat training 8 pat validation, $\eta=0,2$, rete neuronale dopo 400 cicli di apprendimento

Dopo 300 cicli di apprendimento la linea rossa relativa al validation raggiunge il buon valore minimo di $\sum e^2 = 0,8666$.

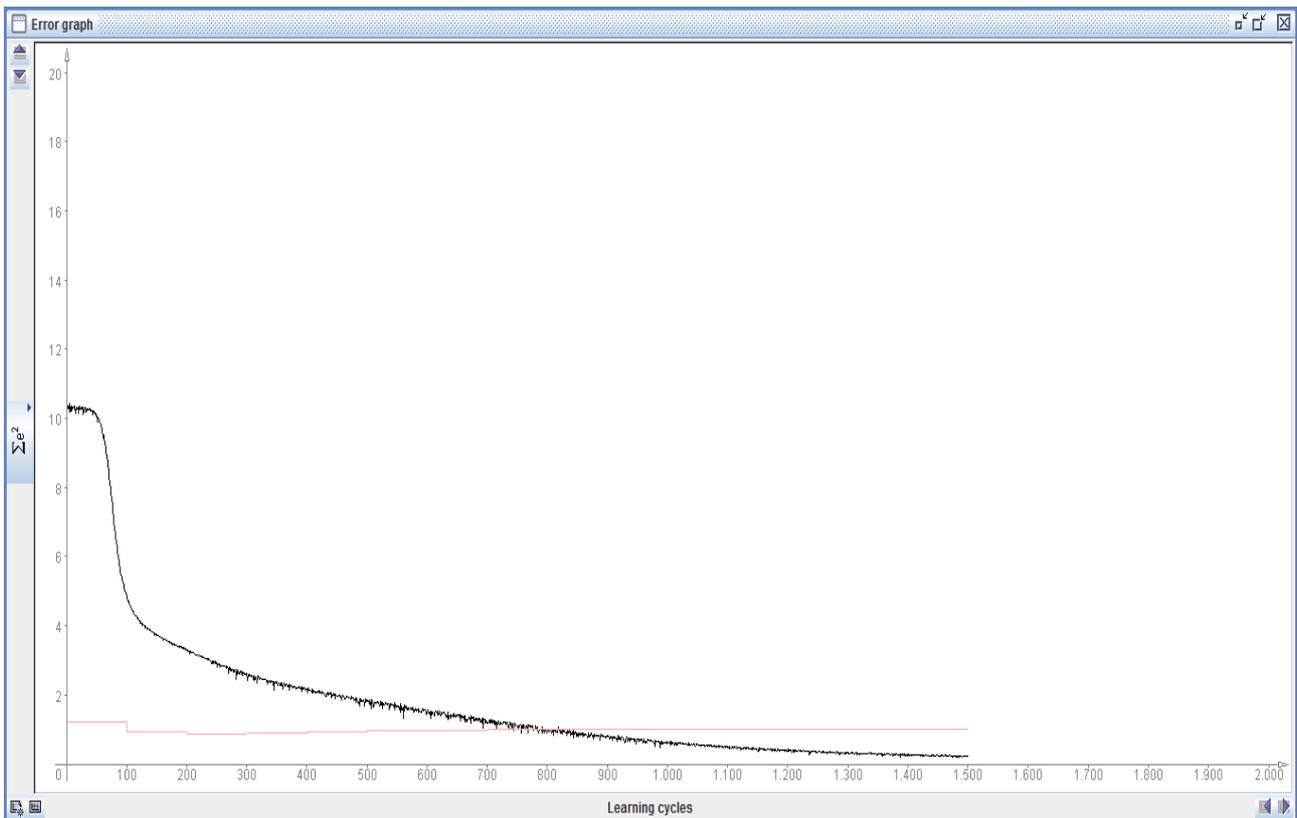


Figura: Armstrong vs Other, 40 pat training 8 pat validation, $\eta=0,2$,
Error Graph della rete neuronale dopo 1500 cicli di apprendimento

Si modifica poi la distribuzione dei pattern nel training e nel validation set, impiegando 16 pattern per autore (quindi 32 in totale) nel file di training set e 8 pattern per autore (quindi 16 in totale) nel validation set.

Si vuole verificare ancora una volta in quale misura il numero e la proporzione di esempi di apprendimento e di test influiscano sui risultati dell'addestramento.

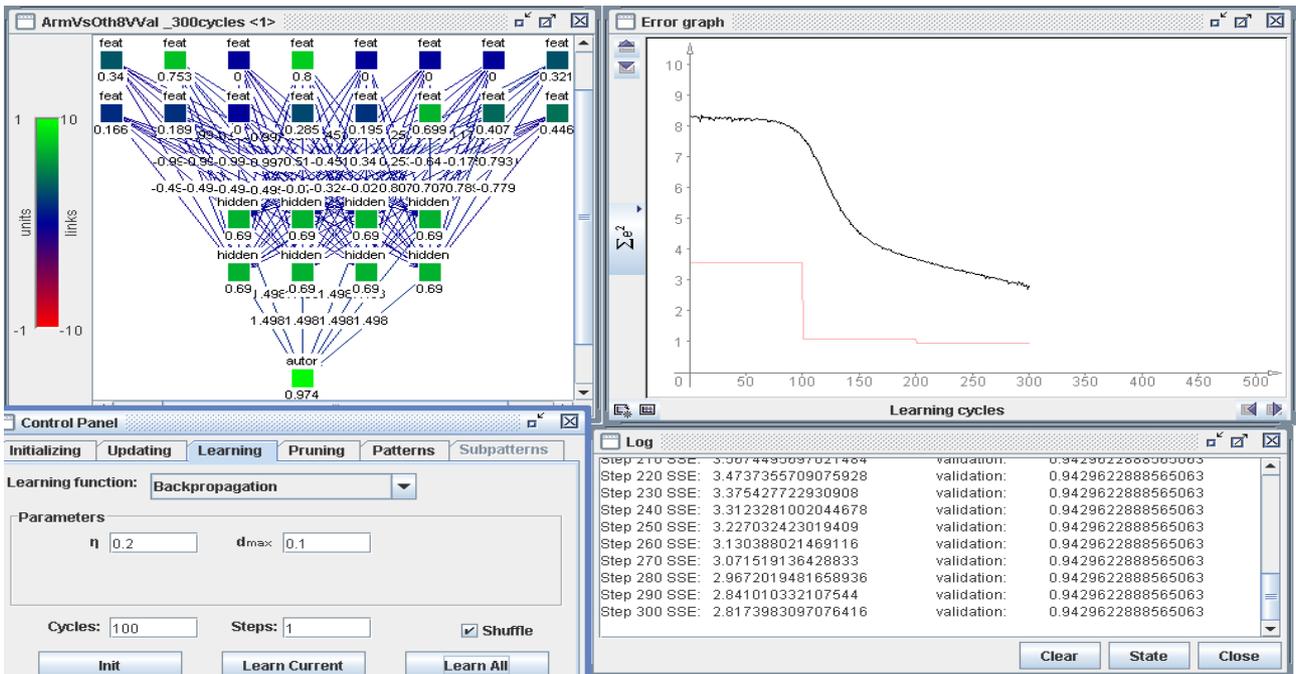


Figura: Armstrong vs Other, 32 pat training 16 pat validation, $\eta=0,2$, rete neuronale dopo 300 cicli di apprendimento

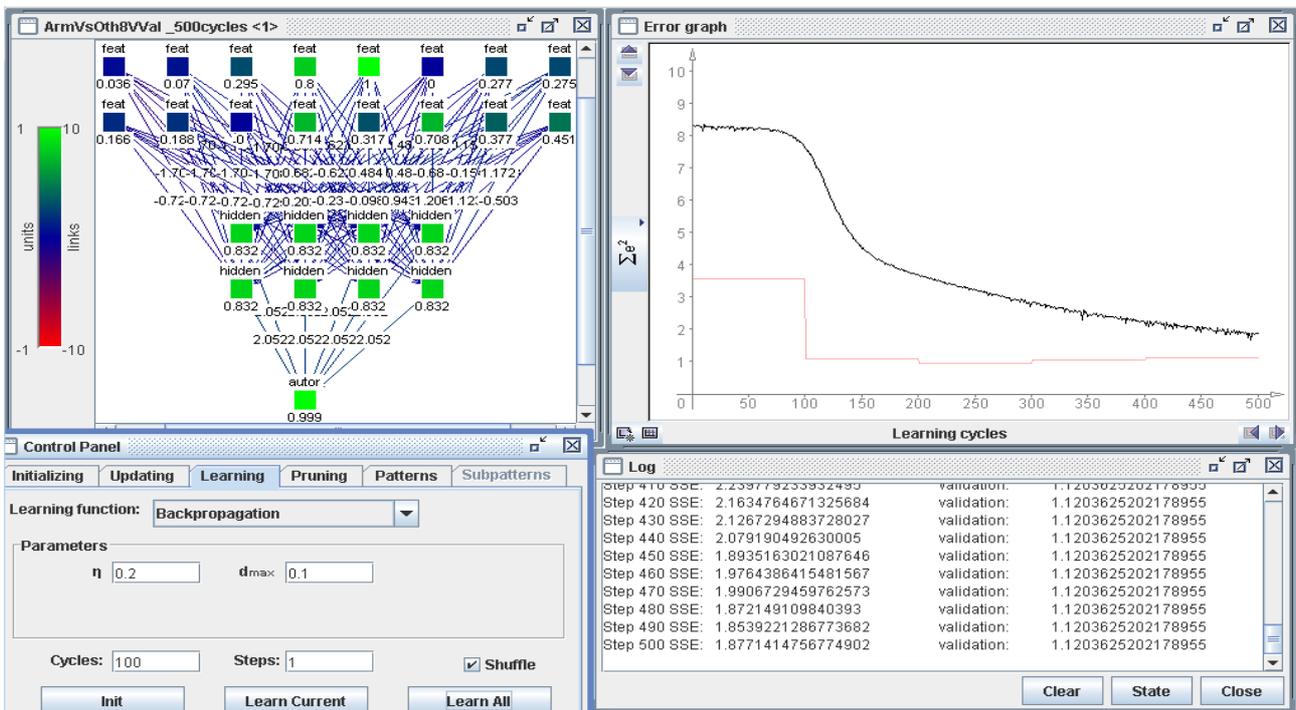


Figura: Armstrong vs Other, 32 pat training 16 pat validation, $\eta=0,2$, rete neuronale dopo 500 cicli di apprendimento

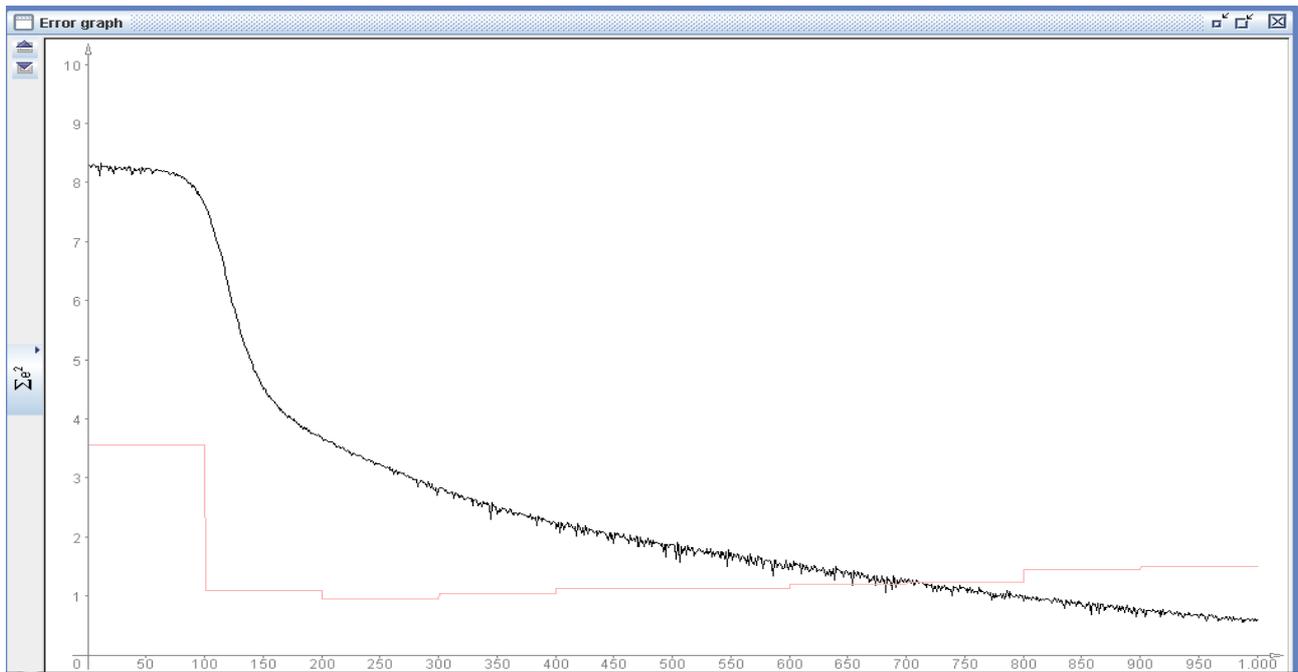


Figura: Armstrong vs Other, 32 pat training 16 pat validation, $\eta=0,2$,
 Error Graph della rete neuronale dopo 1000 cicli di apprendimento

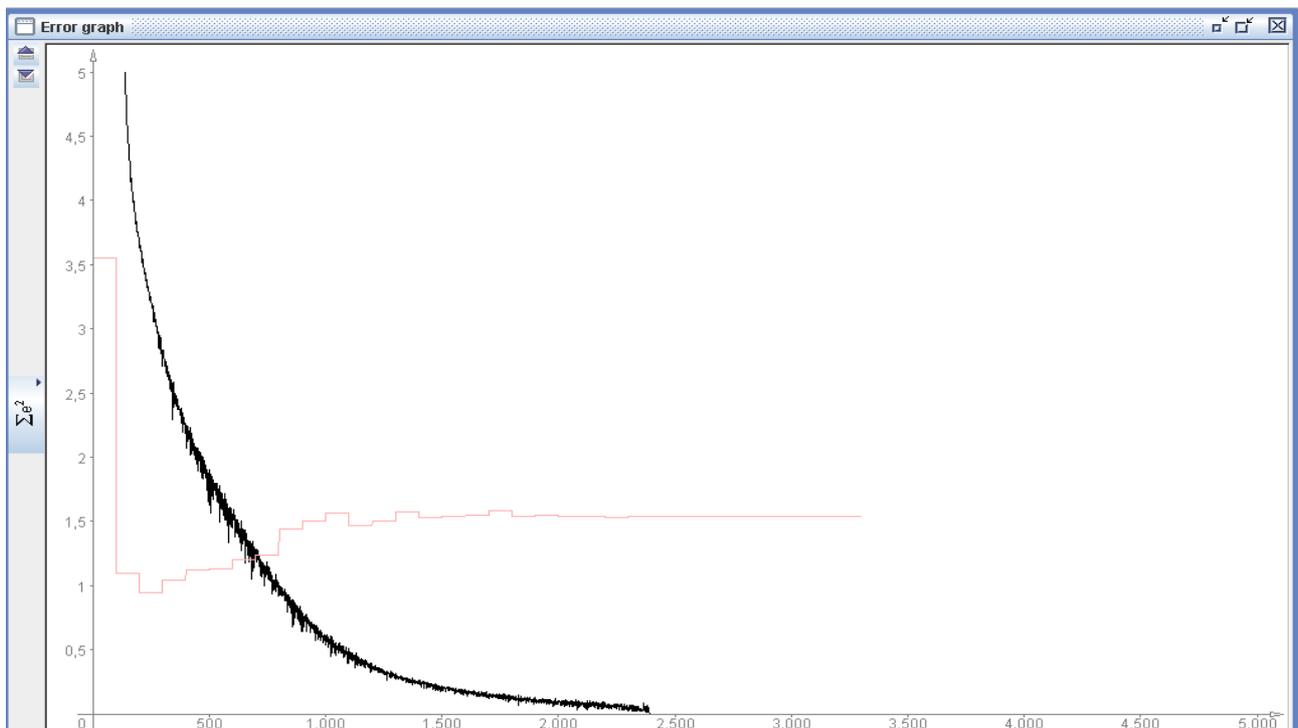


Figura: Armstrong vs Other, 32 pat training 16 pat validation, $\eta=0,2$,
 Error Graph della rete neuronale dopo 3300 cicli di apprendimento

La rete più efficientemente addestrata è quella che si ottiene dopo 300 cicli di apprendimento; la linea rossa relativa al validation set assume il valore $\sum e^2 = 0,9429$.

Si modifica poi nuovamente la distribuzione dei pattern nel training e nel validation set, impiegando 14 pattern per autore (quindi 28 in totale) nel file di training set e 10 pattern per autore (quindi 20 in totale) nel validation set.

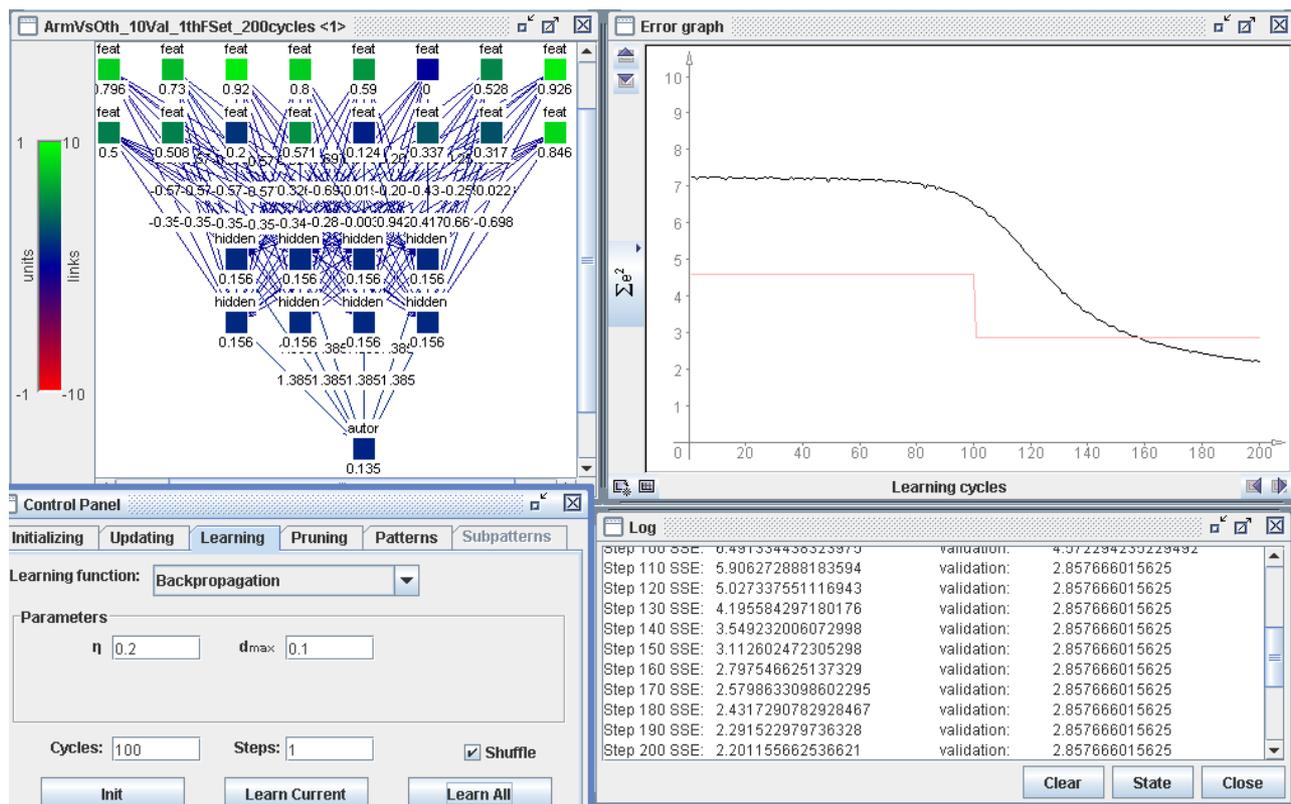


Figura: Armstrong vs Other, 28 pat training 20 pat validation, $\eta=0,2$, rete neuronale dopo 200 cicli di apprendimento

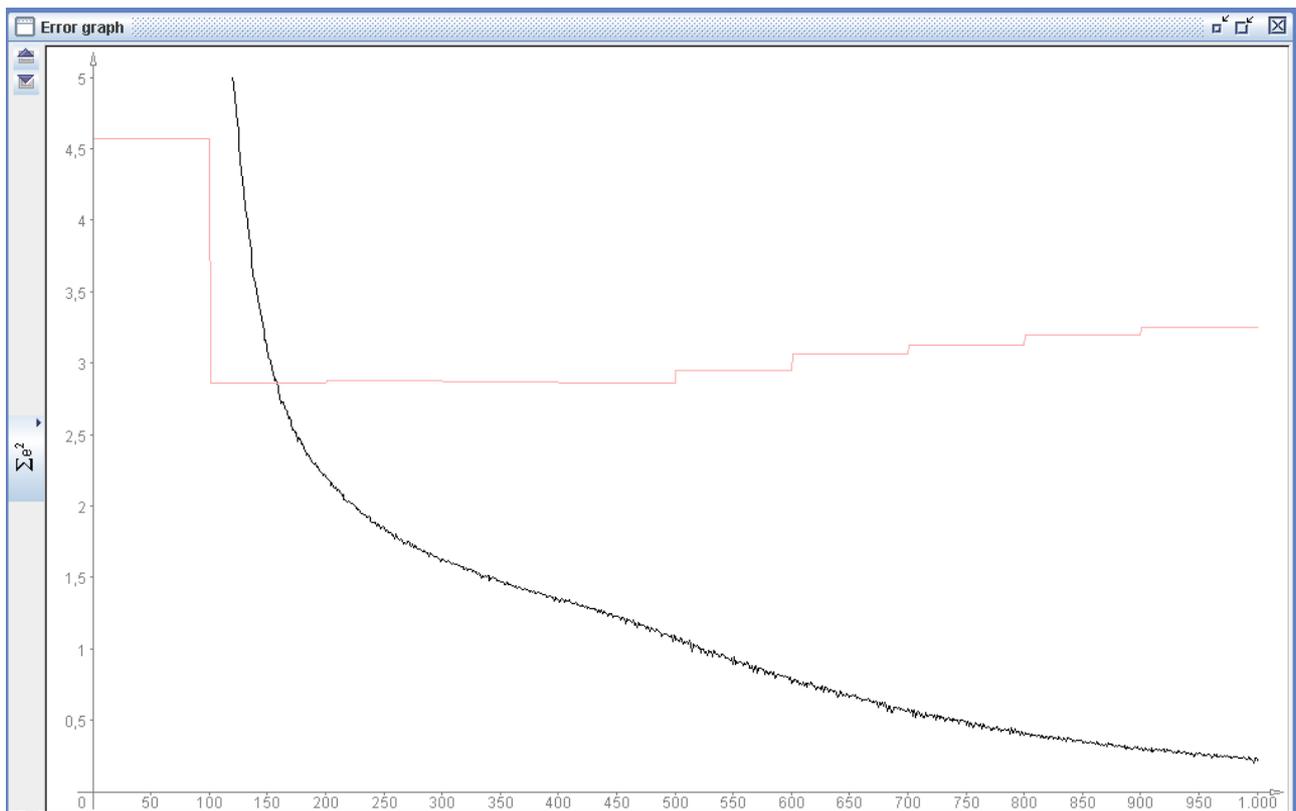


Figura: Armstrong vs Other, 28 pat training 20 pat validation, $\eta=0,2$,
Error Graph della rete neurale dopo 1000 cicli di apprendimento

La rete più efficientemente addestrata è quella che si ottiene dopo 300 cicli di addestramento; la linea rossa relativa al validation set assume il valore $\sum e^2 = 2,8576$.

Questi esperimenti potrebbero portare a concludere che maggiore è il numero dei pattern disponibili nel training set migliori sono le capacità di generalizzazione della rete.

Questo però è vero solo in parte.

Non deve stupire infatti se la somma degli errori al quadrato relativi al validation set è minore nei set che contengono meno pattern, dato che appunto è la somma di meno elementi.

Quindi non è detto che un minimo migliore della linea rossa relativa al validation set sia necessariamente indice di miglior apprendimento nel caso in

cui i validation set contengano quantità diverse di pattern.

Inoltre va detto che un validation set con un numero molto basso di pattern in proporzione al numero totale (ad esempio i soli 8 pattern su 48 del primo di questi tre esperimenti) comporta un test meno oggettivo perchè maggiormente influenzato dai particolari pattern di cui è dotato.

Curiosamente si nota che con lo stesso set di 16 feature, le stesse impostazioni iniziali nel *Control Panel* e la stessa distribuzione dei pattern (32 nel training set e 16 nel validation set) **l'esperimento in cui si confronta Chet Baker con gli altri trombettisti e quello in cui si confronta Luis Armstrong con gli stessi portano a risultati molto diversi.**

Infatti non solo con Armstrong si ottiene un livello di apprendimento molto più alto ($\sum e^2 = 0,9429$), ma con Baker si riesce a portare la linea rossa relativa al validation set ad un minimo solo intervenendo su η , ottenendo comunque valori che si aggirano attorno a $\sum e^2 = 3,8$ dopo svariati cicli di apprendimento.

Si è portati a concludere che le 16 feature scelte evidenziano più marcatamente le caratteristiche di Armstrong piuttosto che quelle di Baker.

Un altro fattore che sicuramente influenza i risultati quello che si può definire lo stile "medio" degli altri trombettisti (B. Berigan, B. Beiderbecke, B. Clayton, B. Hackett, C. Shavers, R. Eldridge e W. Vache) scelti per i due confronti. Esso evidentemente è maggiormente assimilabile a quello di Baker che a quello di Armstrong se si considerano le caratteristiche analizzate dalle 16 feature del set.

Non si sarebbe potuti arrivare a queste deduzioni se ci si fosse limitati al confronto diretto tra i due improvvisatori.

6.3

Beatles e Rolling Stones

In questo paragrafo si confrontano i due gruppi inglesi più famosi della storia del pop-rock nati entrambi nel 1960. Sicuramente **il pop ed il rock si caratterizzano per una minore complessità melodico-armonica e ritmica rispetto al jazz e alla musica classica**. Questo rende **più arduo il processo di codificazione di feature distintive**.

I due gruppi sono inoltre appartenenti allo stesso periodo e anche se chiunque li saprebbe distinguere grazie allo stile inconfondibile ed originale di entrambi, il loro "*sound*" è per diversi aspetti molto simile.

Le difficoltà di discernimento in questi generi quando si trattano informazioni MIDI aumentano ulteriormente poichè **non può essere codificato il timbro delle voci dei cantanti** che è forse il principale aspetto che distingue un gruppo pop-rock da un altro. Nei file MIDI la traccia della voce è eseguita il più delle volte da uno strumento e non sono presenti nemmeno le parole del testo.

E' necessario dunque trovare altri aspetti che caratterizzano le due band.

Ecco l'elenco dei 24 midifile dei Beatles scelti (trovati in questo buon sito <http://freemidipromidi.com/rock/beatles/>):

- Come Together;
- Drive My Car;
- Get Back;
- Help;
- Helter Skelter;
- I'm a Loser;
- Let It Be;

- Love Me Do;
- Lucy in the Sky with Diamonds;
- Michelle;
- ObLaDiObLaDà;
- Please Mister Postman;
- PS I Love You;
- Roll Over Beethoven;
- Sgt Pepper;
- Taxman;
- Tell Me Why;
- Ticket to Ride;
- Twist and Shout;
- While My Guitar Gently Weeps;
- Whit a Little Help From My Friends;
- Yellow SubMarine;
- Yesterday;
- You can't Do That.

Segue l'elenco dei 24 midifile dei Rolling Stones scelti (trovati nello stesso sito

<http://freemidipromidi.com/rock/rollingstones/>) :

- Angie;
- Beast of Burden;
- Brown Sugar;
- Citadel;
- Get off My Cloud;
- Honky Tonk Woman;
- Miss You;

- Mothers Little Helper;
- Out of Time;
- Ruby Tuesday;
- Time is on my Side;
- Under my Thumb;
- Waiting on a Friend;
- Wild Horses;
- You can't Always Get What You Want;
- 19th Nervous Breakdown;
- Gimme Shelter;
- Jumping Jack Flash;
- Let's Spend the Night Together;
- Mixed Emotions;
- Paint it Black;
- Satisfaction;
- Start Me Up;
- Sympathy for the Devil.

Seguendo lo stesso procedimento descritto negli altri paragrafi si sono scelte 20 feature tra le 160 disponibili escludendone alcune a monte prima dell'estrazione dal midifile con JSymbolic e tutte le altre in seguito ad un confronto accurato con valori alla mano, mediando tra evidenze sperimentali e valutazioni teoriche.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		Average Note Duration	Average Number of Independent Voices	Brass Fraction	Electric Instrument Fraction	Glissando Prevalence	Harmonic Complexity	Melodic Complexity	Melodic Complexity	Melodic Complexity	Most Common Melody	Most Common Melody	Note Density	Number of Instruments	Orchestra Pitch Class	Repeat Rate	Saxophone	Strongest Instrument	Variability	Woodwinds	
2	RollingStones	1.351957	4.252000	0.000000	1.058021	0.000000	1.000000	0.698601	0.561016	0.438325	0.696396	0.621348	4.873879	3.805595	0.000000	3.261938	0.619877	0.000000	11.688612	0.372662	0.000000
3	RollingStones	1.164706	4.978000	0.555416	2.082198	0.000000	1.000000	0.698195	0.668437	0.173159	0.719940	0.874082	7.377417	9.000000	0.000000	3.261938	0.530114	0.555416	11.688612	0.174449	0.747863
4	RollingStones	0.757464	4.447000	0.770464	1.411553	0.311271	1.000000	1.011767	0.991435	0.258295	0.803081	1.187147	11.669584	3.261938	0.182973	3.261938	0.573524	0.547431	11.688612	0.875990	0.000000
5	RollingStones	1.752578	4.025000	0.000000	2.275702	0.000000	1.000000	0.727666	0.892130	0.202056	0.846491	1.286107	4.145380	2.718282	0.000000	9.000000	0.846491	0.000000	16.309691	0.821843	0.000000
6	RollingStones	0.726930	2.862000	0.477328	1.780169	0.120883	1.000000	0.265528	0.824050	0.336859	0.916020	0.907559	5.349579	5.000000	0.000000	9.000000	0.916020	0.000000	17.125176	0.816342	0.000000
7	RollingStones	1.136747	6.778000	0.711728	2.348542	0.207063	1.000000	0.265798	0.932207	0.330308	0.914180	0.998057	9.285651	2.990110	0.000000	2.718282	0.914180	0.711728	16.309691	0.396574	0.000000
8	RollingStones	1.048089	4.052000	0.737036	1.311858	0.077767	2.000000	1.111373	0.622820	0.144631	0.850169	1.102167	7.198010	9.000000	0.000000	3.261938	0.438144	0.461775	8.275473	0.940309	0.000000
9	RollingStones	0.627970	3.304000	0.000000	1.097016	0.000000	1.000000	1.158605	0.608105	0.239055	0.956487	1.002104	10.788861	3.533766	0.000000	2.990110	0.956487	0.000000	11.416784	0.662602	0.523125
10	RollingStones	1.567534	4.626000	1.156305	0.893211	0.054318	2.000000	0.375295	0.501420	0.375295	1.442087	0.938828	6.455919	2.718282	0.000000	8.000000	1.442087	0.000000	16.853447	0.105169	0.000000
11	RollingStones	1.795620	4.594000	0.000000	0.000000	0.132234	1.000000	0.627008	0.678370	0.391675	0.602954	0.891740	4.789613	2.718282	0.570438	8.000000	0.602954	0.000000	13.591409	0.580882	0.757832
12	RollingStones	1.555762	4.681000	0.000000	2.501580	0.365541	1.000000	0.427930	1.325102	0.153200	0.810806	0.794987	6.722311	2.718282	0.000000	3.261938	0.491119	0.000000	11.416784	1.253340	0.000000
13	RollingStones	0.804920	2.654000	0.000000	0.466839	0.748269	1.000000	0.000000	0.642685	0.000000	0.943979	1.041835	3.830059	2.718282	0.680209	2.718282	0.943979	0.000000	13.047753	0.430051	0.000000
14	RollingStones	1.279853	4.683000	0.553927	1.088187	0.289212	1.000000	0.894941	0.549980	0.347270	0.605162	0.742013	6.213992	3.261938	0.000000	2.990110	0.605162	0.553927	11.416784	0.451756	1.079434
15	RollingStones	2.673380	4.594000	0.000000	0.857078	0.716600	1.000000	0.875078	0.604426	0.160024	0.737598	0.781376	4.588460	2.990110	0.000000	8.000000	0.737598	0.000000	18.755145	0.755992	0.000000
16	RollingStones	1.104742	4.461000	0.061885	1.029225	0.000000	2.000000	0.226822	0.582353	0.143185	0.785423	1.106581	9.679802	8.000000	0.000000	2.990110	0.785423	0.000000	14.135066	0.465538	0.000000
17	RollingStones	0.882543	5.163000	0.000000	2.127079	0.000000	3.000000	0.613255	1.206108	0.376042	0.642685	1.275070	9.948911	4.077423	0.000000	9.000000	0.579778	0.000000	13.047753	0.856943	0.000000
18	RollingStones	0.984813	5.295000	0.693864	2.294832	0.000000	2.000000	0.438144	0.386641	0.232157	0.666230	1.191194	10.438202	9.000000	0.217221	2.990110	0.666230	0.693864	8.571305	1.196364	0.000000
19	RollingStones	0.958326	4.638000	0.000000	1.151463	0.173635	1.000000	1.103253	0.547772	0.063262	0.645628	1.130861	11.011760	6.000000	0.000000	2.718282	0.645628	0.000000	14.905500	1.301384	0.389360
20	RollingStones	1.017922	3.310000	0.394735	0.776594	0.000000	2.000000	0.949377	1.182154	0.102711	0.780002	1.253733	5.368607	8.000000	0.000000	2.718282	0.678002	0.394735	9.236200	0.741773	0.000000
21	RollingStones	0.547405	1.874000	0.000000	1.211059	0.000000	3.000000	0.623556	0.329000	0.168081	1.800770	1.895315	4.419926	2.718282	0.000000	3.261938	1.800770	0.000000	11.416784	1.083359	0.000000
22	RollingStones	0.374501	2.726000	0.000000	1.788262	0.277115	1.000000	0.280279	0.583457	0.117449	0.715158	0.876657	11.044379	3.261938	0.000000	8.000000	0.704489	0.000000	14.135066	1.072532	0.000000
23	RollingStones	1.107317	4.535000	1.329804	1.371455	0.000000	1.000000	0.711111	0.669774	0.368474	0.711111	11.201494	8.597925	8.000000	0.484771	2.718282	0.711111	1.329804	12.775925	0.765273	0.000000
24	RollingStones	1.057286	5.413000	0.188116	0.439248	0.326023	1.000000	0.686421	0.674691	0.366881	0.629810	1.053239	10.198993	3.533766	0.000000	3.261938	0.629810	0.000000	11.688612	0.764780	0.000000
25	RollingStones	0.911605	3.302000	0.000000	1.930265	0.093948	2.000000	0.465000	0.765998	0.502229	1.027119	0.898362	6.420582	5.000000	0.000000	2.718282	1.027119	0.000000	8.866867	0.747800	0.000000
26	Beatles	1.272863	3.609000	0.000000	2.467000	0.000000	1.000000	0.466907	1.265385	0.398297	1.366672	1.399413	4.055676	7.000000	0.000000	2.718282	1.366672	0.000000	12.339724	0.417175	0.000000
27	Beatles	1.019762	6.125000	0.000000	1.168017	0.000000	1.000000	0.857890	0.967132	0.197256	0.913445	0.729137	10.729058	2.990110	0.000000	3.261938	0.913445	0.000000	9.531882	0.892942	0.000000
28	Beatles	0.886222	3.754000	0.000000	2.083669	0.000000	1.000000	1.274723	0.377076	0.271212	1.168017	1.468207	7.431783	2.990110	0.000000	2.990110	1.168017	0.000000	11.688612	0.445134	0.000000
29	Beatles	0.732080	3.798000	0.918927	0.000000	0.660030	3.000000	0.638918	0.916756	0.391774	0.817796	0.753785	9.690675	2.718282	0.887529	8.000000	0.805656	0.000000	13.047753	0.959993	0.000000
30	Beatles	1.350485	3.813000	0.000000	1.161041	0.000000	1.000000	0.449577	0.961693	0.078166	2.179686	2.030695	4.316632	8.000000	0.000000	3.261938	2.179686	0.000000	24.192708	0.392895	0.000000
31	Beatles	1.038892	3.593000	0.000000	0.975248	0.000000	1.000000	0.668827	1.227982	0.046412	0.805656	0.850537	8.108635	4.349251	0.000000	2.990110	0.517928	0.000000	11.416784	1.182424	0.000000
32	Beatles	2.098752	3.287000	0.279603	1.003575	0.000000	1.000000	0.925558	0.369351	0.092007	1.038156	0.982238	3.490274	6.000000	0.000000	9.000000	0.528275	0.000000	18.484316	0.488176	0.000000
33	Beatles	1.112835	3.015000	0.000000	0.487072	0.000000	1.000000	0.875484	0.382595	0.217619	0.932574	1.649204	6.420582	2.718282	0.000000	2.718282	0.932574	0.000000	11.748599	0.418279	0.000000
34	Beatles	1.578203	3.432000	0.000000	0.442559	0.337556	1.000000	0.492344	0.577571	0.307684	1.123136	1.147416	9.007000	2.718282	0.000000	2.718282	1.123136	0.000000	16.309691	0.586768	0.660711
35	Beatles	2.036213	6.045000	0.000000	0.395839	0.000000	1.000000	0.257947	0.812278	0.782103	0.671380	0.948025	4.971737	3.261938	0.000000	3.261938	0.571205	0.000000	10.873127	0.538943	0.791170
36	Beatles	1.086348	4.054000	0.614152	0.790205	0.000000	1.000000	0.522259	1.024176	0.129546	0.657401	0.886589	6.355343	5.000000	0.000000	2.718282	1.173763	0.000000	12.775925	0.544462	0.000000
37	Beatles	0.831408	2.956000	0.253077	1.410082	0.334819	1.000000	0.148166	0.715893	0.118543	0.088187	0.799402	4.901062	2.990110	0.000000	3.261938	1.088187	0.253077	14.135066	1.238453	0.000000
38	Beatles	1.183100	2.888000	0.000000	0.838765	0.000000	1.000000	0.164838	1.299718	0.326603	0.985917	1.125711	5.436564	8.000000	0.000000	2.990110	0.985917	0.000000	14.406894	0.384066	0.000000
39	Beatles	0.420118	2.443000	0.000000	2.483554	0.195464	1.000000	0.185815	1.007254	0.256403	0.970834	1.108053	9.497677	8.000000	0.000000	2.718282	0.970834	0.000000	11.444955	0.910671	0.000000
40	Beatles	1.202966	3.197000	0.428212	1.328413	0.000000	1.000000	0.938550	0.806392	0.227093	1.238650	0.875921	4.235083	8.000000	0.000000	3.261938	1.238650	0.000000	26.095066	0.398046	0.730241
41	Beatles	1.421854	3.250000	0.439984	1.835351	0.000000	1.000000														

- **Harmonicity of Two Strongest Rhythmic Pulses**, rapporto tra i valori dei due *beat* con picchi più alti di frequenza e il più basso (in termini *bin*);
- **Melodic Octaves**, quantità di intervalli melodici che sono ottave;
- **Melodic Thirds**, quantità di intervalli melodici che sono terze maggiori o minori;
- **Melodic Tritones**, quantità di intervalli melodici che sono tritoni;
- **Most Common Melodic Interval Prevalence**, quantità di intervalli melodici che appartengono all'insieme dell'intervallo più comune;
- **Most Common Pitch Class Prevalence**, quantità di eventi Note On appartenenti alla più comune *classe di nota*;
- **Note Density**, numero medio di note per secondo;
- **Number of Moderate Pulses**, numero di picchi con frequenze normalizzate superiori a 0,01;
- **Orchestral Strings Fraction**, quantità di eventi Note On appartenenti a *patch* di strumenti orchestrali ad arco, ossia le *patch* GeneralMIDI 41 o 47;
- **Pitch Class Variety**, numero di *classi di nota* usate almeno una volta;
- **Repeated Notes**, quantità di note ripetute melodicamente;
- **Saxophone Fraction**, quantità di eventi Note On appartenenti a *patch* di sassofono, ossia le *patch* GeneralMIDI 65 o 68;
- **Strongest Rhythmic Pulse**, valore del *bin* del *beat* con frequenza più alta;
- **Variability of Time Between Attacks**, deviazione standard del tempo trascorso tra gli eventi di Note On considerando tutti i canali.

Non è facile notare delle differenze significative circa i valori delle feature limitandosi ad osservarli nella tabella, come avviene, in misura variabile da caso a caso, negli esperimenti dei precedenti paragrafi.

Dunque si è basata la scelta anche sull'ascolto dei brani stessi.

Ci si è soffermati in particolare sulla **presenza o meno di certi strumenti**

particolari che conferiscono ai brani caratteri distintivi. Lo dimostrano le feature: "Brass Fraction", "Electric Instrument Fraction", "Orchestral Strings Fraction" e "Saxophone Fraction".

Altre feature scelte si concentrano sulla misura di **quanto il ritmo dei brani sia serrato** e sulla loro velocità, come: "Average Note Duration", "Note Density" e "Variability of Time Between Attacks".

Altre ancora analizzano le **differenze dinamiche e la distribuzione delle frequenze**, come: "Harmonicity of Two Strongest Rhythmic Pulses", "Number of Moderate Pulses" e "Strongest Rhythmic Pulse".

Le restanti misurano la **frequenza di certi intervalli e la ripetitività melodico-armonica** ("Most Common Melodic Interval Prevalence", "Most Common Pitch Class Prevalence" e "Repeated Notes").

Scelte le 20 feature giudicate più utili per discernere i due gruppi si passa alla costruzione della rete. Essa presenta 20 ingressi, due livelli da 5 unit nello strato hidden e una unit nello strato di output.

Si impiegano 18 pattern per autore (quindi 36 in totale) nel file di training set e 6 pattern per autore (quindi 12 in totale) nel validation set.

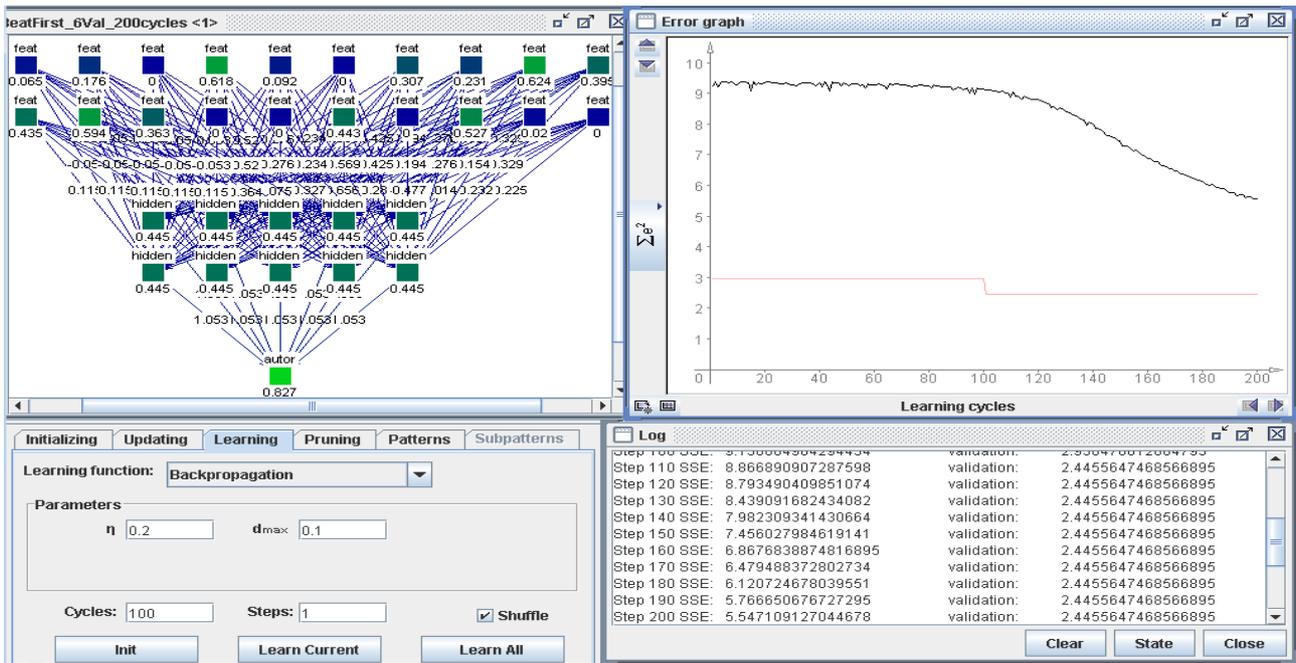


Figura: Beatles vs Rolling Stones, 36 pat training 12 pat validation, $\eta=0,2$, rete neuronale dopo 200 cicli di apprendimento

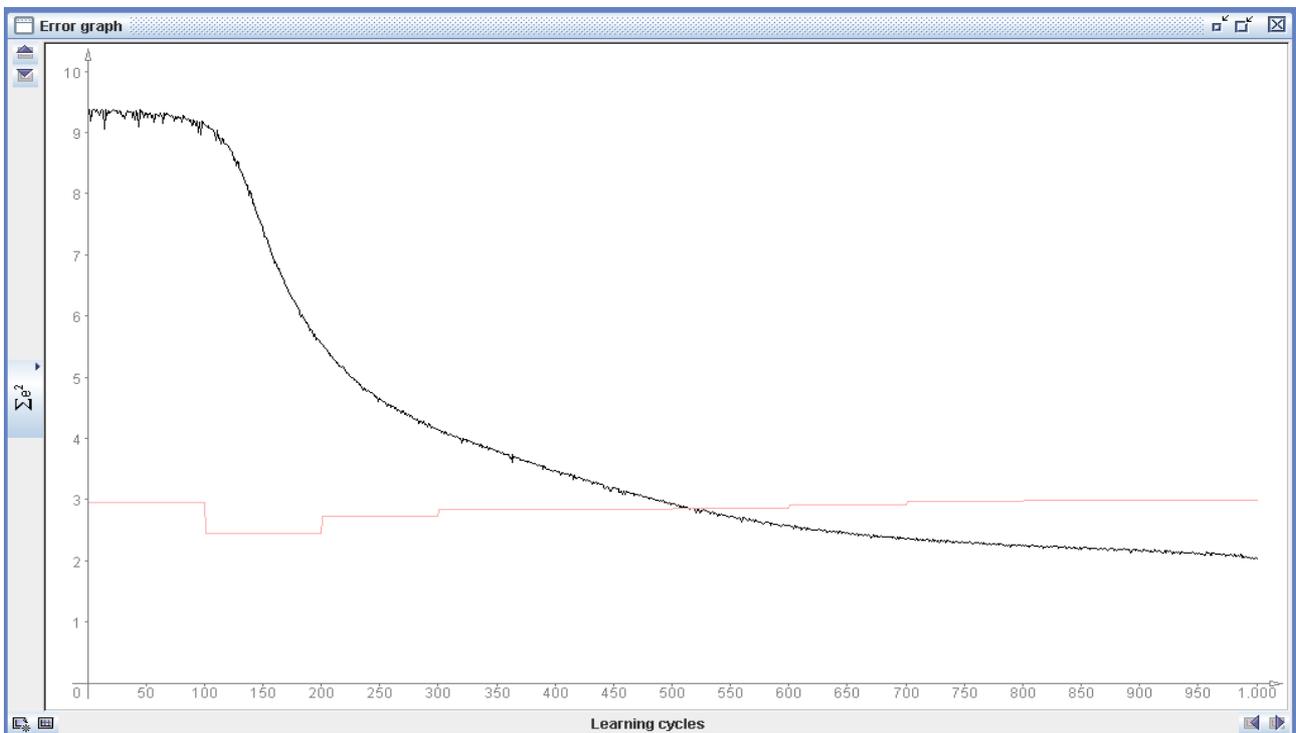


Figura: Beatles vs Rolling Stones, 36 pat training 12 pat validation, $\eta=0,2$, Error Graph della rete neuronale dopo 1000 cicli di apprendimento

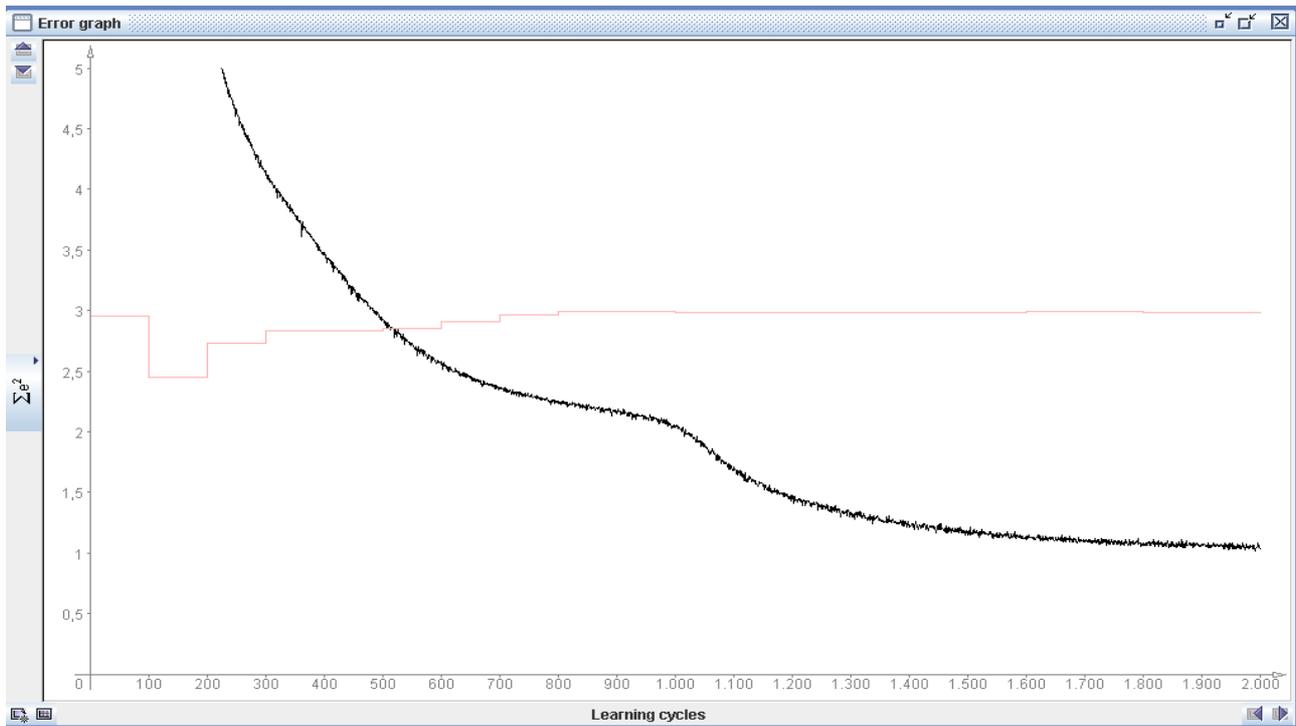


Figura: Beatles vs Rolling Stones, 36 pat training 12 pat validation, $\eta=0,2$,
Error Graph della rete neuronale dopo 2000 cicli di apprendimento

Si nota che la linea rossa relativa al validation set raggiunge dopo 200 cicli il minimo valore di $\sum e^2 = 2,4455$, quindi la rete con il miglior livello di generalizzazione è quella ottenuta dopo 200 cicli di apprendimento.

Si diminuisce poi il parametro η , portandolo a $\eta=0,1$ per verificare eventuali miglioramenti nei risultati dell' algoritmo di Backpropagation.

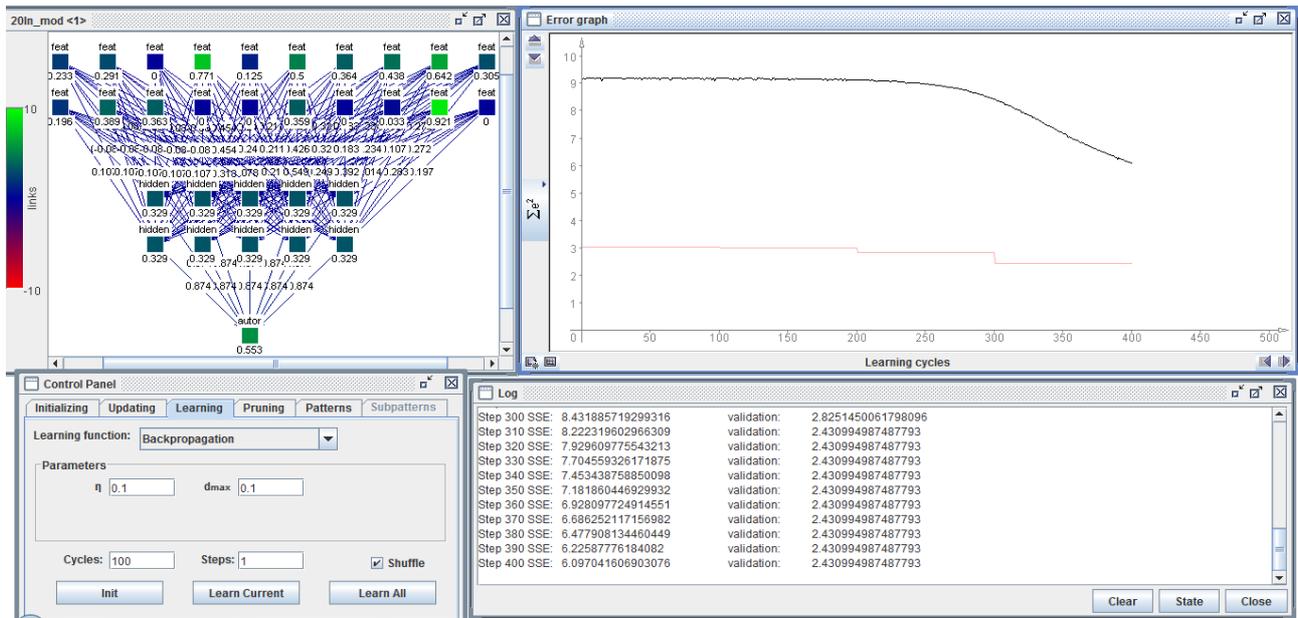


Figura: Beatles vs Rolling Stones, 36 pat training 12 pat validation, $\eta=0,1$,
rete neuronale dopo 400 cicli di apprendimento

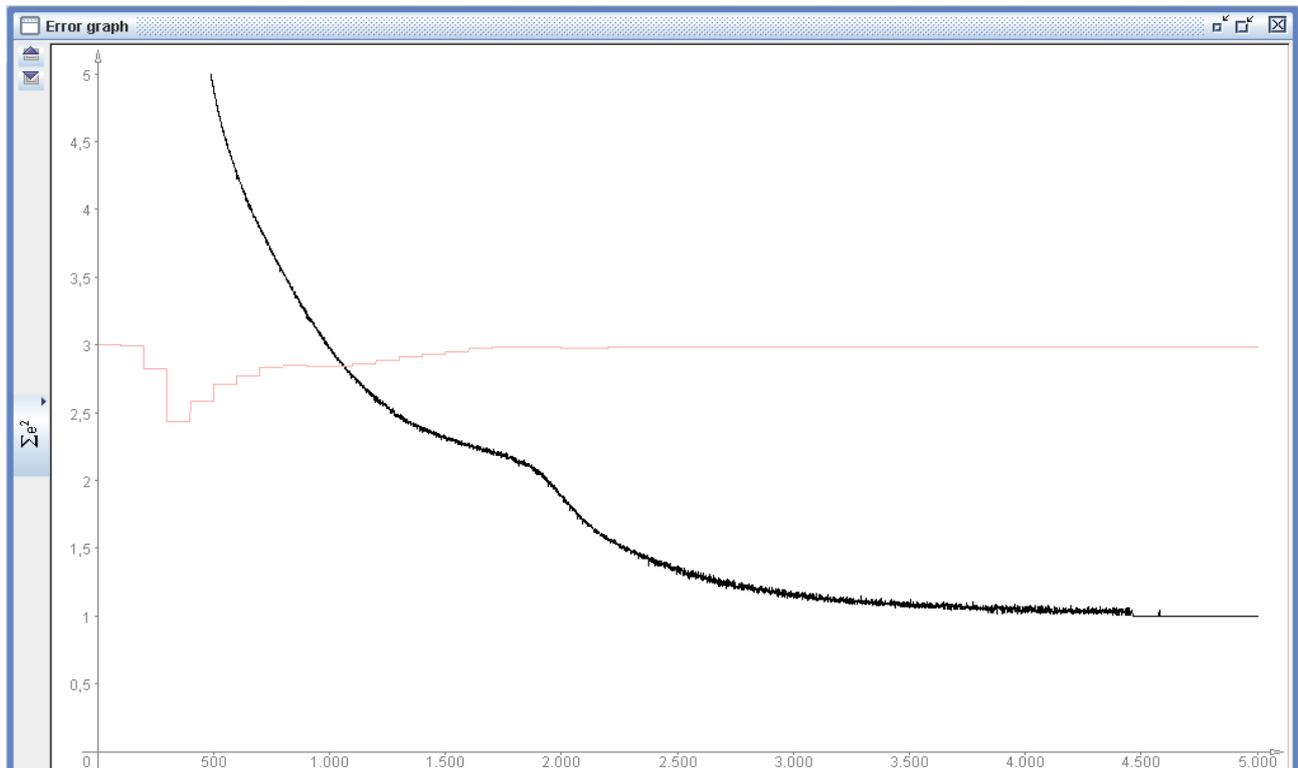


Figura: Beatles vs Rolling Stones, 36 pat training 12 pat validation, $\eta=0,1$,
Error Graph della rete neuronale dopo 5000 cicli di apprendimento

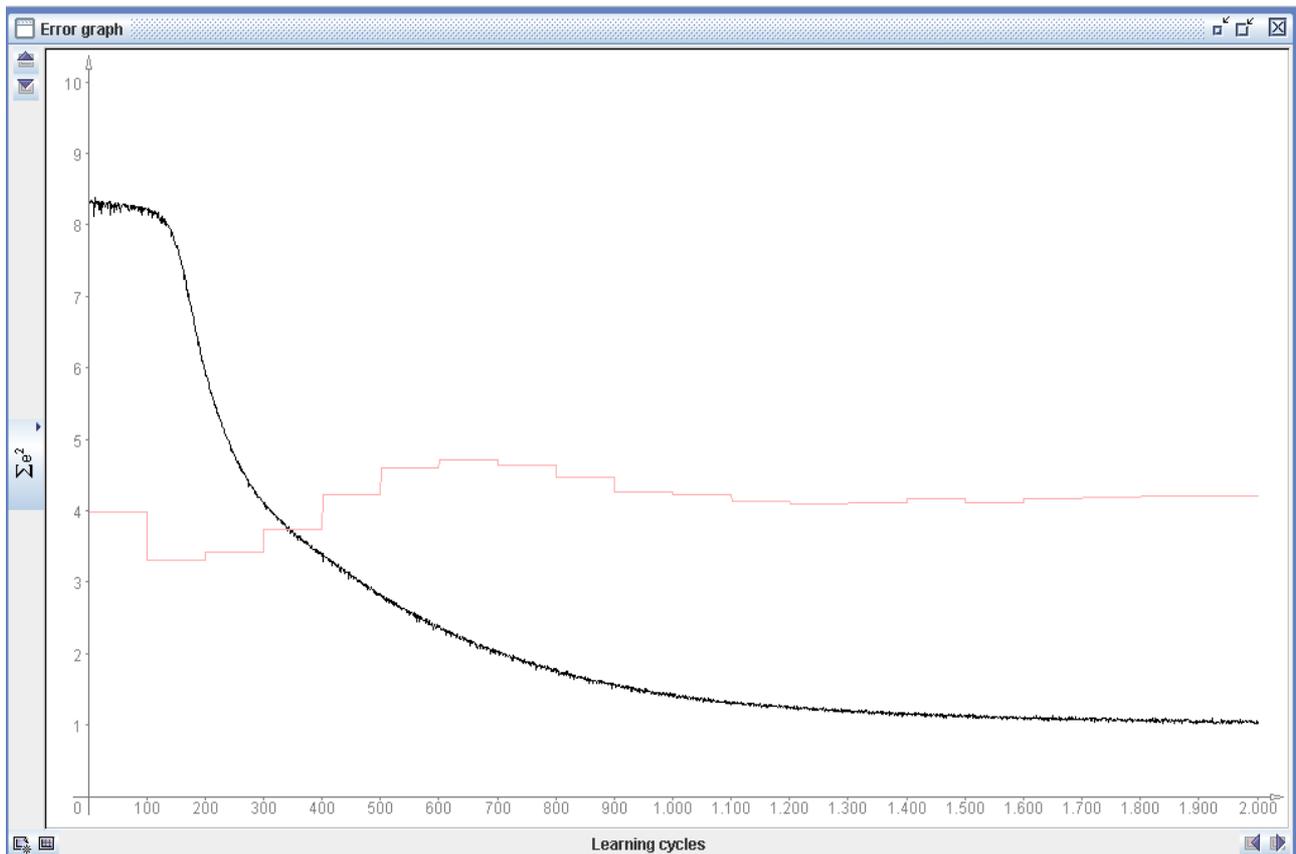


Figura: Beatles vs Rolling Stones, 32 pat training 16 pat validation, $\eta=0,2$,
 Error Graph della rete neuronale dopo 2000 cicli di apprendimento

Si nota che la linea rossa relativa al validation set raggiunge dopo 200 cicli il minimo valore di $\sum e^2 = 3,3050$, quindi la rete con il miglior livello di generalizzazione è quella ottenuta dopo 200 cicli di apprendimento.

6.4

Beatles e Led Zeppelin

In questo paragrafo si confrontano due band; la prima è quella protagonista anche del paragrafo precedente, i Beatles, la seconda sono i **Led Zeppelin**. Questi ultimi si formano più tardi dei Beatles, nel 1968 per la precisione, e sono caratterizzati da uno **stile più aggressivo** definito da molti Hard-Blues; sono tra i padri indiscussi dell'Hard Rock, genere che ha molto successo negli anni successivi.

Si intende verificare se questa più marcata differenza di genere rispetto alle due band confrontate nel paragrafo precedente, comporti anche una più facile classificazione tramite reti neurali.

Va comunque specificato che le due band hanno sì stili piuttosto differenti, ma presentano anche molti aspetti in comune che rendono ostica l'estrapolazione di caratteristiche distintive. Fra tutti si sottolinea l'ampio uso di **scale blues e pentatoniche**, la predominanza di strumenti quali **la chitarra elettrica, il basso e la batteria** e le canzoni quasi sempre in metri di **4/4**, così come del resto accade per la maggior parte delle band pop-rock e blues.

I brani MIDI utilizzati sono gli stessi del capitolo precedente per quanto riguarda i Beatles ed i seguenti 24, scaricati dal medesimo sito, per quanto riguarda i Led Zeppelin:

- Battle of Evermore;
- Black Country Woman;
- Black Mountain Side;
- Celebration Day;
- Communication Breakdown;
- Custard Pie;

- Dancing Days;
- Dazed and Confused;
- Down By the Seaside;
- Dyer Maker;
- Gallow's Pole;
- Good Times Bad Times;
- Heartbreaker;
- Houses of the Holy;
- Immigrant Song;
- In My Time of Dying;
- In the Light;
- Living Loving Maid;
- Moby Dick;
- No Quarter;
- Ramble On;
- Rock and Roll;
- Since I've Been Loving You;
- Whole Lotta Love.

Si decide di confrontare le due band **suddividendo le feature** messe a disposizione da JSymbolic in **tre gruppi**: quelle di **tipo armonico-melodico**, quelle **riguardanti gli strumenti utilizzati e le dinamiche** e **quelle di tipo ritmico**.

Per il primo gruppo di tipo armonico-melodico si selezionano 20 feature.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Amount of	Average	Direction	Interval	Melodic	Melodic	Melodic	Melodic	Most Cor	Most Comp	Pitch Vari	Range	Range of	Relative	Relative	Repeted	Relative	Strong T	Voice Eq	Stepwise	
2	Beatles	2.250000	0.000000	1.890000	7.000000	1.080000	0.470000	1.270000	0.400000	0.000000	2.000000	9.790000	14.950000	1.400000	1.610000	1.830000	1.370000	0.076800	3.000000	2.400000	0.790000
3	Beatles	2.340000	0.000000	1.900000	4.620000	1.340000	0.860000	0.700000	0.200000	0.000000	2.000000	12.500000	16.310000	2.640000	0.680000	2.710000	0.910000	0.081300	2.000000	3.570000	0.800000
4	Beatles	2.540000	0.000000	2.260000	3.260000	0.580000	1.270000	0.380000	0.270000	0.000000	9.000000	11.420000	16.310000	2.580000	1.470000	1.840000	1.170000	0.182100	1.000000	2.490000	0.410000
5	Beatles	2.080000	2.890000	0.840000	5.000000	0.490000	0.640000	0.920000	0.390000	5.000000	5.000000	10.600000	15.490000	0.900000	1.030000	3.620000	0.810000	0.184800	3.000000	1.840000	0.650000
6	Beatles	2.590000	0.000000	1.580000	6.520000	0.230000	0.450000	0.960000	0.080000	0.000000	4.000000	11.690000	16.310000	0.740000	2.940000	1.200000	2.180000	0.092400	1.000000	6.360000	0.880000
7	Beatles	2.000000	0.000000	2.480000	2.720000	1.330000	0.670000	1.230000	0.050000	4.000000	7.000000	9.240000	12.500000	0.290000	1.840000	2.950000	0.520000	0.088700	3.000000	2.660000	1.220000
8	Beatles	1.410000	0.000000	1.680000	4.000000	0.530000	0.930000	0.370000	0.090000	2.000000	0.000000	10.600000	16.310000	2.210000	1.230000	1.870000	0.530000	0.135900	2.000000	4.470000	1.240000
9	Beatles	2.250000	0.000000	1.600000	3.260000	0.520000	0.880000	0.380000	0.220000	0.000000	7.000000	7.880000	12.500000	1.280000	1.680000	2.050000	0.930000	0.083500	3.000000	2.490000	0.620000
10	Beatles	2.110000	9.670000	1.720000	6.520000	1.170000	0.490000	0.580000	0.310000	0.000000	2.000000	9.790000	14.680000	0.950000	1.570000	2.740000	1.120000	0.236500	1.000000	1.700000	1.110000
11	Beatles	1.470000	0.000000	2.490000	7.000000	1.280000	0.260000	0.810000	0.780000	5.000000	0.000000	11.960000	12.780000	1.250000	1.570000	2.900000	0.580000	0.266400	2.000000	2.000000	0.920000
12	Beatles	1.890000	0.000000	1.760000	5.000000	0.860000	0.520000	1.020000	0.130000	3.000000	5.000000	12.500000	25.550000	1.100000	0.470000	2.710000	1.170000	0.078300	2.000000	3.670000	0.700000
13	Beatles	2.000000	12.740000	1.870000	2.000000	0.520000	0.150000	0.720000	0.120000	0.000000	9.000000	9.790000	13.050000	0.690000	1.530000	2.460000	1.090000	0.092400	2.000000	1.410000	0.890000
14	Beatles	2.600000	0.000000	1.530000	4.000000	0.630000	0.160000	1.300000	0.330000	0.000000	0.000000	7.610000	11.140000	1.530000	1.700000	3.320000	0.990000	0.231100	2.000000	1.730000	0.480000
15	Beatles	2.590000	0.000000	1.780000	5.000000	0.490000	0.190000	1.010000	0.260000	0.000000	2.000000	7.610000	12.500000	1.600000	1.600000	2.850000	0.970000	0.252800	2.000000	3.080000	0.580000
16	Beatles	2.630000	0.000000	1.950000	3.000000	0.940000	0.940000	0.810000	0.230000	0.000000	7.000000	11.960000	13.050000	2.220000	0.770000	1.280000	1.240000	0.176700	2.000000	1.760000	0.520000
17	Beatles	2.530000	0.000000	1.990000	6.000000	1.210000	0.430000	0.850000	0.510000	2.000000	2.000000	9.240000	15.490000	1.360000	1.550000	3.250000	0.610000	0.100600	3.000000	1.730000	0.860000
18	Beatles	1.770000	0.000000	1.000000	5.000000	0.830000	0.510000	0.710000	0.370000	9.000000	2.000000	8.150000	13.050000	0.920000	1.300000	3.250000	0.540000	0.231100	2.000000	2.930000	0.790000
19	Beatles	2.090000	0.000000	2.420000	7.000000	0.620000	0.250000	1.000000	0.200000	5.000000	4.000000	10.870000	15.770000	1.400000	1.400000	3.150000	0.380000	0.083500	2.000000	1.870000	0.470000
20	Beatles	2.200000	0.000000	2.270000	3.260000	0.820000	0.390000	0.560000	0.490000	0.000000	9.000000	8.700000	11.140000	1.970000	1.700000	1.800000	1.180000	0.222900	3.000000	1.260000	0.940000
21	Beatles	1.980000	1.520000	0.860000	7.000000	0.860000	0.230000	1.190000	0.490000	7.000000	4.000000	10.870000	13.860000	2.890000	2.890000	2.740000	0.440000	0.079800	2.000000	2.480000	0.810000
22	Beatles	2.010000	0.000000	1.840000	7.000000	1.330000	0.390000	0.370000	0.000000	0.000000	4.000000	10.330000	14.140000	1.130000	1.840000	2.190000	1.010000	0.074600	2.000000	2.880000	0.750000
23	Beatles	1.770000	0.000000	2.130000	2.000000	0.440000	0.330000	0.540000	0.090000	0.000000	7.000000	10.600000	14.410000	0.900000	1.320000	2.560000	0.990000	0.089400	2.000000	3.670000	1.140000
24	Beatles	1.460000	0.000000	1.420000	2.000000	1.030000	0.250000	1.150000	0.290000	2.000000	5.000000	9.240000	18.760000	2.930000	0.800000	3.350000	0.470000	0.077600	2.000000	5.760000	0.790000
25	Beatles	2.450000	0.000000	1.650000	4.000000	0.470000	0.370000	0.930000	0.330000	0.000000	7.000000	11.960000	14.140000	0.990000	0.990000	1.480000	1.270000	0.222900	3.000000	6.740000	0.700000
26	LedZepplin	2.140000	0.000000	0.940000	5.160000	0.790000	0.350000	0.380000	0.560000	0.000000	9.000000	15.220000	17.130000	1.400000	1.170000	1.620000	1.250000	0.263700	2.000000	5.080000	0.620000
27	LedZepplin	2.640000	6.280000	1.020000	3.260000	0.380000	0.210000	0.960000	0.070000	0.000000	4.000000	13.590000	17.670000	0.510000	1.750000	0.820000	1.720000	0.093100	1.000000	2.030000	0.410000
28	LedZepplin	2.040000	13.490000	1.490000	7.000000	1.060000	0.730000	0.610000	0.650000	2.000000	9.000000	11.140000	13.860000	0.510000	2.090000	3.560000	0.530000	0.082800	2.000000	2.530000	0.760000
29	LedZepplin	1.610000	0.000000	0.880000	5.000000	0.590000	0.140000	1.030000	0.190000	1.000000	0.000000	11.690000	13.050000	0.850000	1.070000	3.390000	0.660000	0.001200	3.000000	2.630000	1.130000
30	LedZepplin	1.890000	14.220000	1.660000	2.720000	1.320000	0.810000	0.420000	0.050000	2.000000	4.000000	12.780000	15.770000	1.210000	0.950000	3.430000	0.710000	0.077600	2.000000	1.270000	1.320000
31	LedZepplin	1.830000	0.000000	1.540000	3.260000	0.940000	1.000000	0.610000	0.130000	2.000000	0.000000	14.140000	17.400000	0.420000	0.420000	3.630000	0.440000	0.266400	2.000000	3.310000	0.540000
32	LedZepplin	2.450000	16.550000	1.840000	9.000000	0.820000	0.260000	0.850000	0.170000	0.000000	7.000000	8.700000	14.950000	1.340000	1.270000	2.310000	0.870000	0.073900	2.000000	3.030000	0.690000
33	LedZepplin	2.040000	0.000000	1.250000	5.000000	0.590000	0.530000	0.370000	0.090000	0.000000	9.000000	10.870000	15.490000	1.360000	1.360000	3.030000	0.830000	0.269100	2.000000	3.990000	0.800000
34	LedZepplin	1.990000	13.800000	2.060000	2.000000	0.450000	0.300000	0.610000	0.200000	0.000000	4.000000	14.140000	18.210000	1.980000	1.040000	3.200000	0.790000	0.079800	1.000000	1.990000	0.760000
35	LedZepplin	1.960000	10.800000	1.970000	5.160000	0.670000	0.350000	0.780000	0.490000	0.000000	9.000000	14.950000	15.220000	0.920000	0.920000	3.180000	0.740000	0.088700	2.000000	1.890000	1.150000
36	LedZepplin	2.060000	0.000000	1.720000	5.000000	0.370000	0.740000	0.950000	0.150000	0.000000	9.000000	9.240000	13.590000	0.000000	2.800000	2.940000	0.740000	0.263700	2.000000	2.400000	0.650000
37	LedZepplin	2.870000	2.820000	1.930000	7.000000	0.840000	0.600000	0.730000	0.220000	0.000000	6.000000	11.960000	13.050000	1.690000	1.690000	2.630000	1.170000	0.266400	2.000000	2.060000	0.480000
38	LedZepplin	1.840000	12.550000	1.470000	6.520000	0.450000	0.150000	0.560000	0.280000	2.000000	9.000000	12.230000	15.220000	1.250000	1.250000	3.310000	0.660000	0.233800	2.000000	1.000000	1.020000
39	LedZepplin	1.460000	12.000000	1.880000	6.520000	0.360000	0.690000	0.670000	0.140000	2.000000	9.000000	14.410000	16.310000	1.130000	0.980000	2.280000	1.280000	0.081300	1.000000	2.640000	1.640000
40	LedZepplin	1.890000	13.030000	1.980000	6.520000	1.200000	0.190000	0.550000	0.160000	0.000000	9.000000	11.690000	14.140000	0.780000	1.060000	3.430000	0.960000	0.269100	2.000000	1.550000	1.330000
41	LedZepplin	1.630000	4.600000	1.290000	7.000000	0.220000	0.990000	0.640000	0.000000	2.000000	7.000000	10.060000	13.320000	2.930000	1.650000	1.340000	0.580000	0.119600	2.000000	2.010000	1.730000
42	LedZepplin	1.830000	0.000000	1.670000	7.000000	1.070000	1.060000	0.470000	0.300000	2.000000	1.000000	15.220000	20.930000	1.050000	0.880000	3.160000	0.700000	0.073900	2.000000	2.620000	1.140000
43	LedZepplin	1.290000	0.000000	0.950000	5.160000	0.730000	0.440000	1.100000	0.050000	5.160000	4.000000	9.790000	13.050000	0.540000	1.610000	2.990000	0.540000	0.171300	1.000000	7.180000	0.700000
44	LedZepplin	2.220000	8.430000	1.840000	13.050000	1.240000	0.930000	0.430000	0.180000	0.000000	9.000000	14.410000	16.310000	1.900000	2.390000	1					

appartenenti ai due toni MIDI più usati;

- **Melodic Fifths**, quantità di intervalli melodici che sono quinte perfette;
- **Melodic Octaves**, quantità di intervalli melodici che sono ottave;
- **Melodic Thirds**, quantità di intervalli melodici che sono terze minori o maggiori;
- **Melodic Tritones**, quantità di intervalli melodici che sono tritoni;
- **Most Common Melodic Interval**, intervallo melodico con la frequenza più alta;
- **Most Common Pitch Class**, valore del *bin* della *classe di nota* più usata;
- **Pitch Variety**, numero di note usate almeno una volta;
- **Range**, differenza tra la nota più alta e quella più bassa;
- **Range of Highest Line**, differenza tra la nota più alta e quella più bassa eseguite nel canale con il più alto tono medio divisa per la differenza tra la nota più alta e la nota più bassa del pezzo;
- **Relative Range of Loudest Voice**, differenza tra la nota più alta e la nota più bassa eseguite nel canale con il più basso tono medio divisa per la differenza tra la nota più alta e quella più bassa di tutto il pezzo;
- **Relative Strength of Most Common Intervals**, quantità di intervalli melodici che appartengono al secondo intervallo melodico più usato divisa per la quantità di intervalli melodici appartenenti all'intervallo melodico più usato;
- **Repeated Notes**, quantità di note ripetute melodicamente;
- **Relative Strength of Top Pitches**, frequenza del secondo tono più usato divisa per la frequenza del tono più usato;
- **Strong Tonal Centres**, numero di picchi nell'istogramma delle quinte che ammontano ad almeno il 9 % di tutti gli eventi Note On;
- **Voice Equality - Melodic Leaps**, deviazione standard della media dei salti melodici nei toni MIDI per ogni canale che contiene almeno una note;
- **Stepwise Motion**, quantità di intervalli melodici che corrispondono ad una

seconda minore o maggiore.

Si costruisce poi la rete con 20 ingressi, due strati da 5 unit nello strato hidden e una unit nello strato di output.

Si impiegano 18 pattern per autore (quindi 36 in totale) nel file di training set e 6 pattern per autore (quindi 12 in totale) nel validation set.

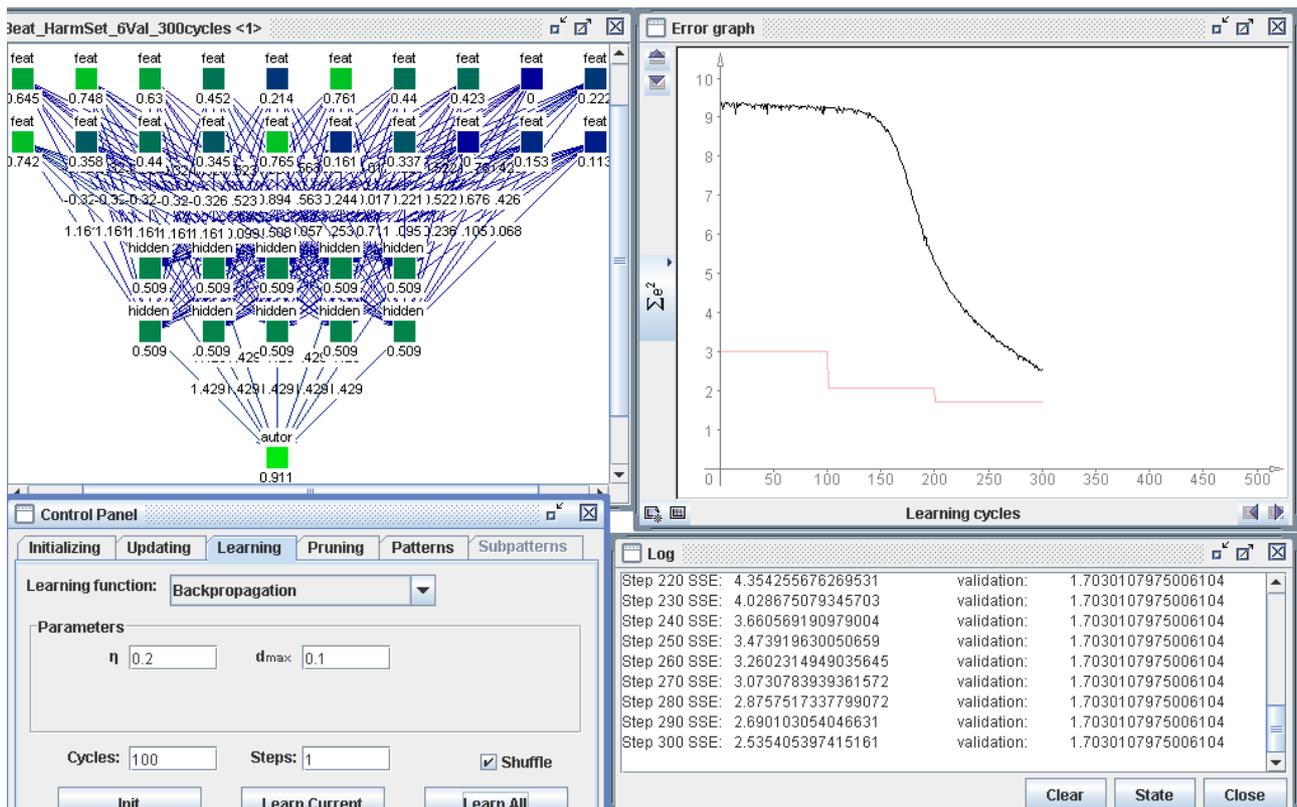


Figura: Beatles vs Led Zepplin, set di feature armonico-melodico,

36 pat training 12 pat validation, $\eta=0,2$,

rete neuronale dopo 300 cicli di apprendimento

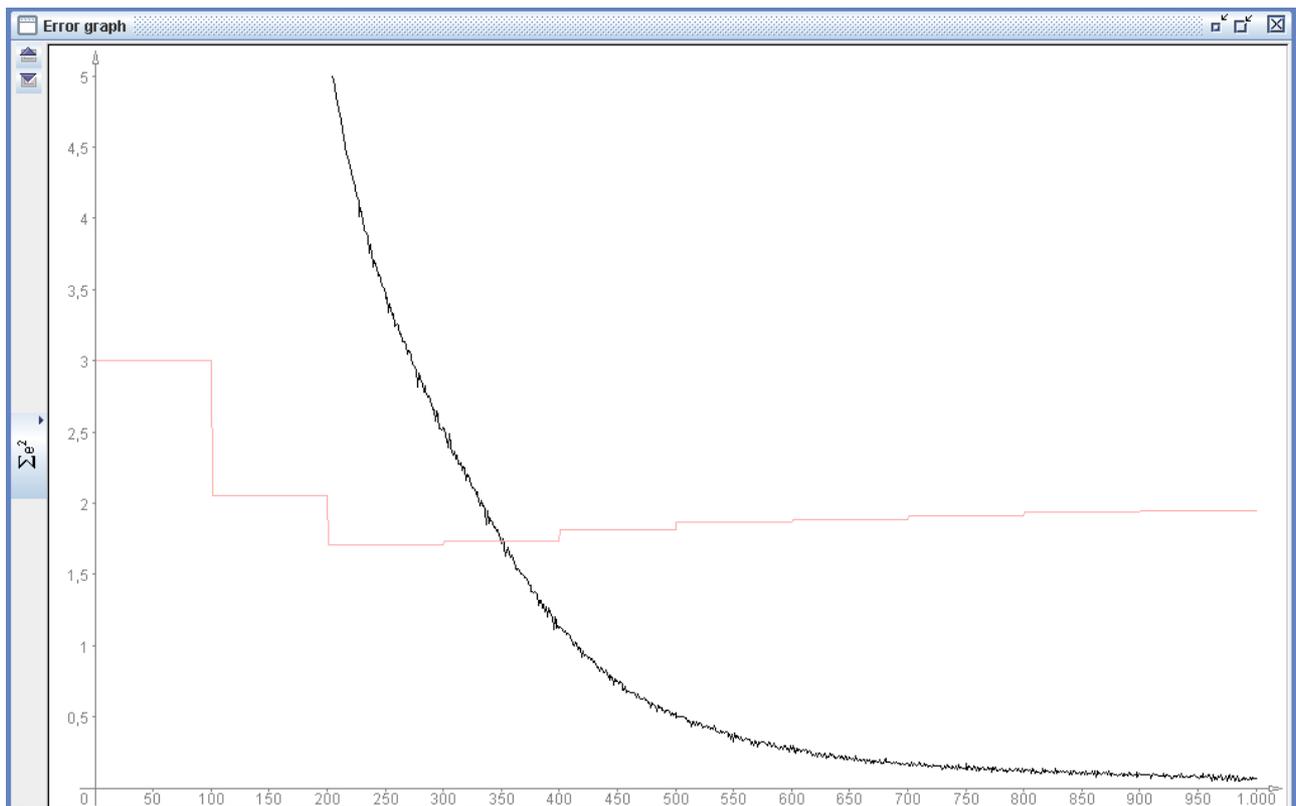


Figura: Beatles vs Led Zeppelin, set di feature armonico-melodico,
 36 pat training 12 pat validation, $\eta=0,2$,
 Error Graph della rete neurale dopo 1000 cicli di apprendimento

Si nota che la linea rossa relativa al validation set raggiunge dopo 300 cicli il minimo valore di $\sum e^2 = 1,7030$, quindi la rete con il miglior livello di generalizzazione è quella ottenuta dopo 300 cicli di apprendimento.

Si modifica poi la distribuzione dei pattern nel training e nel validation set, impiegando 16 pattern per autore (quindi 32 in totale) nel file di training set e 8 pattern per autore (quindi 16 in totale) nel validation set.

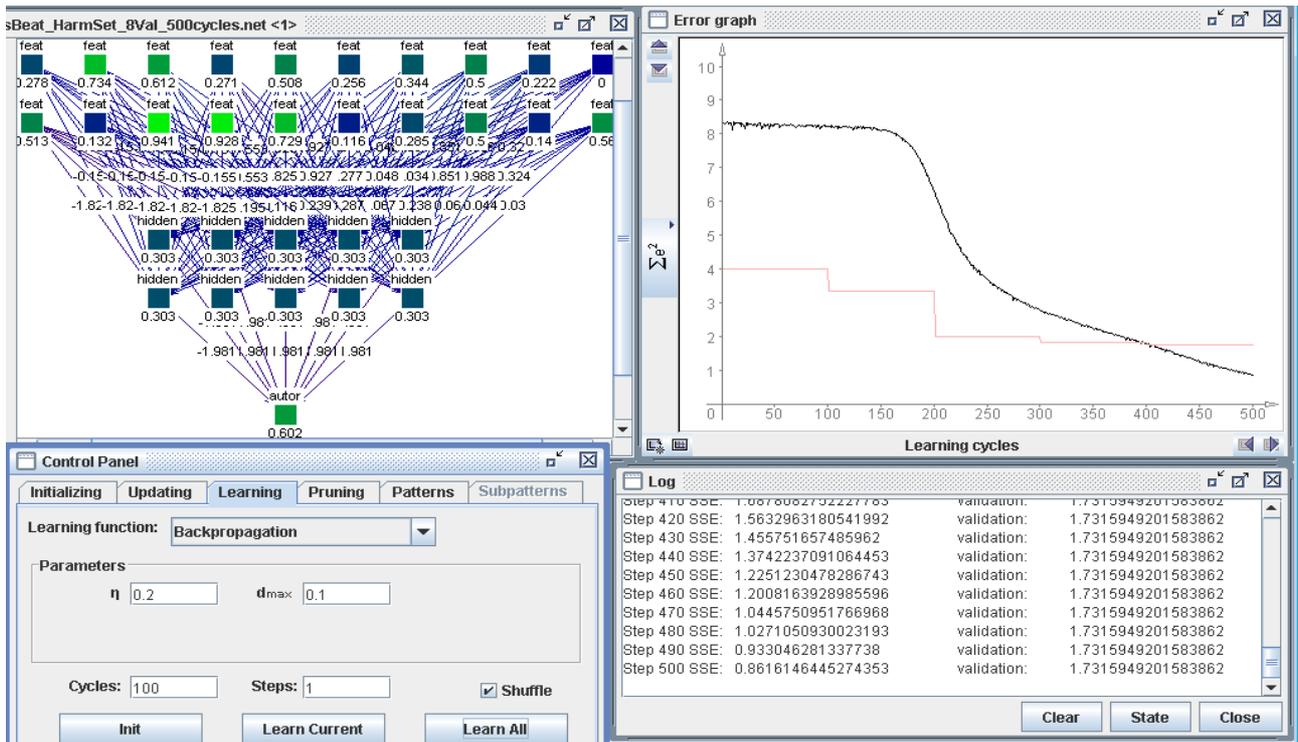


Figura: Beatles vs Led Zepplin, set di feature armonico-melodico,
 32 pat training 16 pat validation, $\eta=0,2$,
 rete neurale dopo 500 cicli di apprendimento

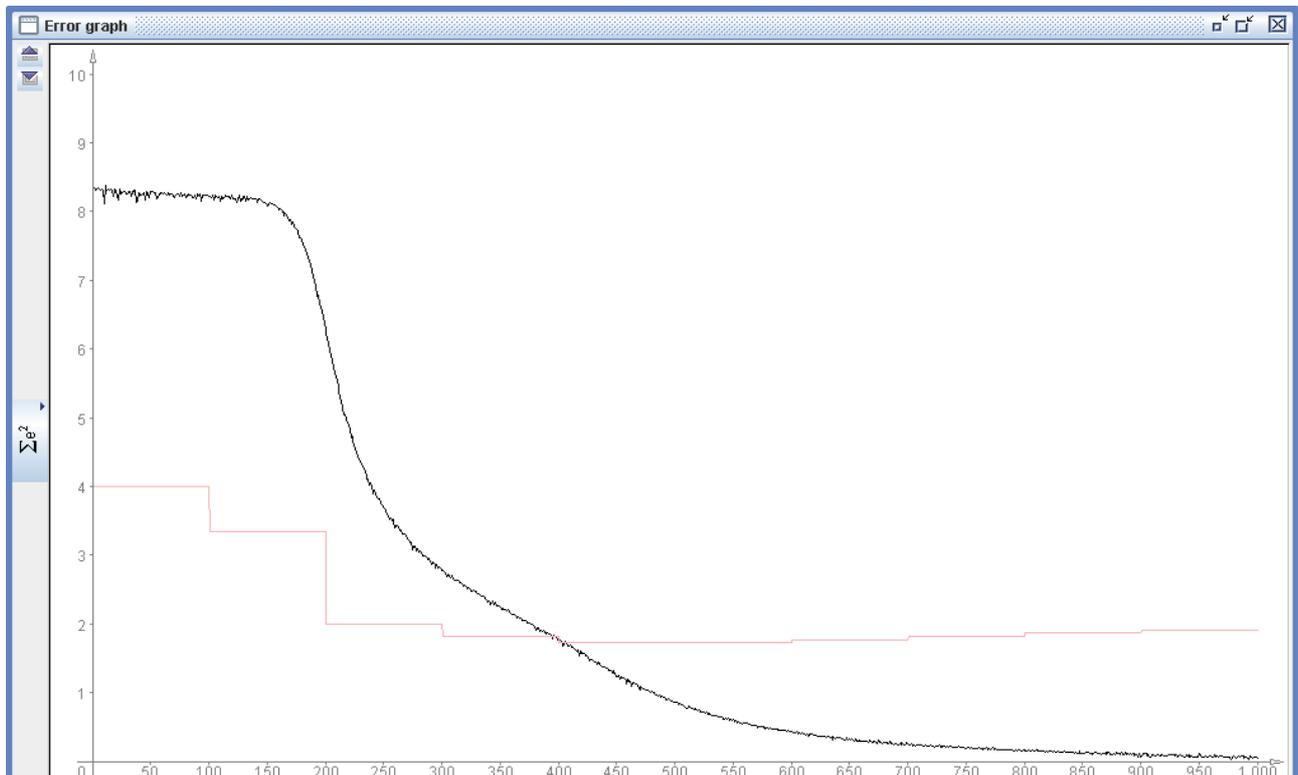


Figura: Beatles vs Led Zepelin, set di feature armonico-melodico,
 32 pat training 16 pat validation, $\eta=0,2$,
 Error Graph della rete neurale dopo 1000 cicli di apprendimento

Si nota che la linea rossa relativa al validation set raggiunge dopo 500 cicli il minimo valore di $\sum e^2 = 1,7315$, quindi la rete con il miglior livello di generalizzazione è quella ottenuta dopo 500 cicli di apprendimento.

Il valore è molto simile a quello della prova precedente ($\sum e^2 = 1,7030$) nonostante la diversa distribuzione di pattern nel training e nel validation set.

Si selezionano poi 20 feature per formare il set riguardante le dinamiche e gli strumenti utilizzati.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U					
1	Acoustic	Average	Average	Brass	Fr	Electric	Electric	Importan	Importan	Importan	Importan	Orchestra	Percussi	String	Er	String	Key	Variability	Variabili	Vibrato	P	Violin	Fr	Voice	S	Woodwinds
2	Beatles	0,000000	1,910000	3,610000	0,000000	1,250000	2,470000	1,790000	0,520000	3,670000	1,370000	0,000000	1,210000	0,000000	1,250000	32,440000	1,330000	0,000000	0,000000	0,000000	7,200000	0,000000	0,000000	0,000000	0,000000	0,000000
3	Beatles	1,160000	8,160000	6,130000	0,000000	0,520000	1,170000	1,770000	0,450000	6,640000	1,460000	0,000000	1,480000	0,000000	0,370000	15,220000	1,620000	0,000000	0,000000	0,000000	4,750000	0,000000	0,000000	0,000000	0,000000	0,000000
4	Beatles	0,000000	9,290000	3,750000	0,000000	1,630000	2,080000	1,790000	0,580000	9,130000	1,300000	0,000000	1,080000	0,000000	1,030000	22,230000	1,070000	0,000000	0,000000	0,000000	8,250000	0,000000	0,000000	0,000000	0,000000	0,000000
5	Beatles	1,600000	9,870000	3,800000	0,920000	0,000000	0,000000	0,870000	0,450000	9,930000	2,360000	0,890000	1,150000	0,290000	0,000000	47,490000	1,080000	0,000000	0,000000	0,000000	7,800000	0,000000	0,000000	0,000000	0,000000	0,000000
6	Beatles	0,000000	2,400000	3,810000	0,000000	1,160000	1,160000	2,560000	0,230000	6,910000	1,060000	0,000000	1,220000	0,000000	0,000000	31,640000	3,610000	0,000000	0,000000	0,000000	4,830000	0,000000	0,000000	0,000000	0,000000	0,000000
7	Beatles	1,820000	3,490000	3,590000	0,000000	0,500000	0,980000	0,850000	0,490000	10,960000	2,700000	0,000000	0,820000	0,000000	0,000000	20,350000	3,380000	0,000000	0,000000	0,000000	8,000000	0,000000	0,000000	0,000000	0,000000	0,000000
8	Beatles	0,000000	1,980000	3,290000	0,280000	0,370000	1,000000	1,180000	0,980000	6,860000	2,230000	0,000000	0,720000	0,000000	1,530000	31,430000	1,350000	0,000000	0,000000	0,000000	4,570000	0,000000	0,000000	0,000000	0,000000	0,000000
9	Beatles	2,240000	2,720000	3,020000	0,000000	0,000000	0,490000	1,280000	1,060000	9,840000	2,110000	0,000000	0,720000	0,000000	0,000000	71,020000	3,540000	0,000000	0,000000	0,000000	8,250000	0,000000	0,000000	0,000000	0,000000	0,000000
10	Beatles	0,000000	1,080000	3,430000	0,000000	0,440000	0,440000	1,530000	1,210000	6,600000	0,940000	0,000000	1,170000	0,000000	0,400000	22,650000	1,360000	0,000000	0,000000	0,000000	5,500000	0,660000	0,000000	0,000000	0,000000	0,000000
11	Beatles	1,880000	1,930000	6,050000	0,000000	0,560000	0,400000	0,870000	0,460000	5,840000	2,680000	0,000000	0,900000	0,000000	0,000000	51,010000	1,940000	0,000000	0,000000	0,000000	4,290000	0,790000	0,000000	0,000000	0,000000	0,000000
12	Beatles	0,000000	9,560000	4,050000	0,610000	0,000000	0,790000	1,170000	0,720000	4,890000	2,320000	0,000000	0,790000	0,570000	1,490000	42,580000	1,650000	0,000000	0,000000	0,000000	7,800000	0,000000	0,000000	0,000000	0,000000	0,000000
13	Beatles	0,000000	9,420000	2,960000	0,250000	0,790000	1,410000	1,170000	0,340000	7,580000	2,410000	0,000000	1,550000	0,300000	0,000000	13,720000	1,440000	0,000000	0,000000	0,000000	4,140000	0,000000	0,000000	0,000000	0,000000	0,000000
14	Beatles	1,230000	3,830000	2,890000	0,000000	0,430000	0,840000	0,940000	0,000000	7,330000	2,740000	0,000000	1,610000	0,000000	0,000000	23,870000	1,000000	0,000000	0,000000	0,000000	3,260000	0,000000	0,000000	0,000000	0,000000	0,000000
15	Beatles	0,000000	4,480000	2,440000	0,000000	1,950000	2,480000	1,030000	0,150000	7,070000	2,610000	0,000000	1,200000	0,000000	0,000000	22,920000	1,040000	0,000000	0,000000	0,000000	6,750000	0,000000	0,000000	0,000000	0,000000	0,000000
16	Beatles	0,000000	2,670000	3,200000	0,430000	0,830000	1,330000	1,710000	0,620000	6,030000	1,800000	0,000000	1,190000	0,000000	0,000000	11,360000	2,170000	0,000000	0,000000	0,000000	6,600000	0,730000	0,000000	0,000000	0,000000	0,000000
17	Beatles	0,000000	0,860000	3,250000	0,440000	1,020000	1,840000	1,240000	0,540000	7,340000	1,900000	0,000000	1,200000	0,000000	0,000000	13,720000	2,870000	0,000000	0,000000	0,000000	5,720000	0,000000	0,000000	0,000000	0,000000	0,000000
18	Beatles	0,000000	5,660000	3,450000	0,410000	1,650000	2,000000	0,580000	0,210000	9,150000	3,040000	0,000000	1,270000	0,000000	0,000000	43,870000	1,280000	0,000000	0,000000	0,000000	6,800000	0,000000	0,000000	0,000000	0,000000	0,000000
19	Beatles	0,000000	5,800000	4,410000	0,000000	1,490000	2,060000	1,070000	0,940000	9,150000	1,670000	0,000000	1,090000	0,000000	0,600000	58,480000	1,300000	0,000000	0,000000	0,000000	7,800000	0,000000	0,000000	0,000000	0,000000	0,000000
20	Beatles	0,820000	2,910000	2,740000	0,000000	1,120000	1,550000	1,650000	0,000000	4,960000	2,030000	0,000000	1,310000	0,000000	0,000000	31,060000	1,080000	0,000000	0,000000	0,000000	6,700000	0,000000	0,000000	0,000000	0,000000	0,000000
21	Beatles	1,710000	1,150000	3,190000	0,000000	0,000000	0,460000	0,960000	0,590000	4,850000	2,830000	0,000000	0,640000	0,000000	1,250000	71,890000	1,250000	0,000000	0,000000	0,000000	4,400000	0,000000	0,000000	0,000000	0,000000	0,000000
22	Beatles	0,000000	5,110000	3,160000	0,000000	0,500000	0,490000	1,960000	0,700000	5,250000	1,530000	0,000000	1,390000	0,000000	1,360000	28,030000	1,070000	0,000000	0,000000	0,000000	5,800000	0,430000	0,000000	0,000000	0,000000	0,000000
23	Beatles	0,000000	5,890000	2,570000	1,440000	0,000000	0,000000	0,940000	0,400000	13,220000	2,340000	0,000000	0,950000	0,000000	0,000000	10,910000	1,450000	0,000000	0,000000	0,000000	3,800000	1,040000	0,000000	0,000000	0,000000	0,000000
24	Beatles	2,380000	3,460000	3,100000	0,000000	0,000000	0,000000	1,240000	0,820000	10,670000	2,220000	0,900000	0,000000	0,670000	0,000000	23,220000	1,060000	0,000000	0,000000	0,000000	6,000000	0,000000	0,000000	0,000000	0,000000	0,000000
25	Beatles	0,480000	7,160000	3,740000	0,390000	1,250000	1,660000	1,370000	1,310000	7,460000	1,960000	0,000000	1,000000	0,000000	0,000000	32,220000	1,370000	0,000000	0,000000	0,000000	6,670000	0,000000	0,000000	0,000000	0,000000	0,000000
26	LedZepelin	0,000000	3,070000	5,530000	0,000000	1,630000	2,090000	0,580000	1,380000	9,080000	1,710000	0,780000	0,620000	0,000000	0,000000	53,870000	1,400000	0,000000	0,000000	0,000000	3,820000	0,000000	0,000000	0,000000	0,000000	0,000000
27	LedZepelin	0,000000	8,220000	4,700000	0,570000	2,170000	2,600000	2,240000	1,020000	10,570000	1,160000	0,000000	0,800000	0,000000	0,000000	53,960000	1,080000	0,000000	0,000000	0,000000	6,290000	0,000000	0,000000	0,000000	0,000000	0,000000
28	LedZepelin	0,000000	0,700000	4,430000	0,000000	1,770000	2,030000	1,980000	0,390000	8,430000	1,590000	0,000000	1,010000	0,000000	1,060000	57,690000	1,630000	0,350000	0,000000	0,000000	5,140000	0,000000	0,000000	0,000000	0,000000	0,000000
29	LedZepelin	0,000000	0,000000	5,160000	0,000000	1,750000	2,040000	0,380000	1,770000	0,000000	1,530000	0,700000	0,840000	0,000000	0,000000	20,750000	1,530000	0,000000	0,000000	0,000000	5,170000	0,000000	0,000000	0,000000	0,000000	0,000000
30	LedZepelin	0,000000	3,400000	3,340000	0,000000	1,110000	1,840000	2,250000	1,150000	8,040000	1,120000	0,000000	1,610000	0,000000	0,000000	52,730000	1,440000	0,310000	0,000000	0,000000	3,380000	0,650000	0,000000	0,000000	0,000000	0,000000
31	LedZepelin	0,630000	0,480000	5,780000	0,490000	1,880000	2,050000	0,480000	0,940000	8,920000	2,270000	0,060000	0,650000	0,000000	0,610000	20,230000	1,290000	0,000000	0,000000	0,000000	3,000000	0,000000	0,000000	0,000000	0,000000	0,000000
32	LedZepelin	0,000000	1,340000	2,980000	0,000000	1,220000	2,240000	0,960000	0,960000	7,720000	2,450000	0,000000	1,110000	0,000000	4,800000	36,010000	1,050000	0,000000	0,000000	0,000000	9,750000	0,000000	0,000000	0,000000	0,000000	0,000000
33	LedZepelin	2,280000	1,090000	4,260000	0,000000	0,000000	0,830000	1,660000	1,330000	9,240000	1,660000	0,000000	0,510000	0,000000	0,000000	36,560000	2,160000	0,000000	0,150000	6,430000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000
34	LedZepelin	0,000000	4,120000	3,720000	0,000000	1,360000	1,850000	1,790000	0,660000	6,450000	1,700000	0,000000	1,450000	0,000000	0,000000	46,730000	1,790000	0,650000	0,000000	5,330000	0,790000	0,000000	0,000000	0,000000	0,000000	0,000000
35	LedZepelin	0,000000	1,500000	2,880000	0,840000	1,610000	1,960000	1,550000	0,540000	12,260000	1,580000	0,000000	1,260000	0,000000	0,000000	43,060000	1,130000	0,000000	0,000000</							

I valori mostrati nella tabella sono relativi alle seguenti 20 feature riguardanti le dinamiche e gli strumenti utilizzati:

- **Acoustic Guitar Fraction**, quantità di eventi Note On appartenenti a *patch* di chitarra acustica, ossia le *patch* GeneralMIDI 25 o 26;
- **Average Note To Note Dynamics Change**, differenza media di intensità, calcolata in MIDI *velocity units*, tra una nota e la seguente nel medesimo canale;
- **Average Number of Independent Voices**, numero medio di canali differenti in cui sono eseguite note simultaneamente, i resti non sono inclusi nel calcolo;
- **Brass Fraction**, quantità di eventi Note On appartenenti a *patch* di ottoni, ossia le *patch* GeneralMIDI dalla 57 alla 68;
- **Electric Guitar Fraction**, quantità di eventi Note On appartenenti a *patch* di chitarra elettrica, ossia le *patch* GeneralMIDI dalla 27 alla 32;
- **Electric Instrument Fraction**, quantità di eventi Note On appartenenti a *patch* di strumenti elettrici, ossia le *patch* GeneralMIDI 5, 6, 17, 19, dalla 27 alla 32 o dalla 34 alla 40;
- **Importance of Bass Register**, quantità di eventi Note On con toni MIDI dallo 0 al 54;
- **Importance of High Register**, quantità di eventi Note On con toni MIDI dallo 73 al 127;
- **Importance of Loudest Voice**, differenza tra l'intensità media del canale con intensità più alta e l'intensità media degli altri canali che contengono almeno una nota;
- **Importance of Middle Register**, quantità di eventi Note On con toni MIDI dallo 55 al 72;
- **Orchestral Strings Fraction**, quantità di eventi Note On appartenenti a *patch* di archi da orchestra, ossia le *patch* GeneralMIDI dalla 41 alla 47;
- **Percussion Prevalence**, numero totale di eventi Note On corrispondenti a

- strumenti non accordati diviso per il numero totale di eventi Note On in tutto il pezzo;
- **String Ensemble Fraction**, quantità di eventi Note On appartenenti a *patch* di ensemble d'archi, ossia le *patch* GeneralMIDI dalla 49 alla 52;
 - **String Keyboard Fraction**, quantità di eventi Note On appartenenti a *patch* di tastiere con suono d'arco, ossia le *patch* GeneralMIDI dalla 41 alla 47;
 - **Variability of Note Prevalence of Pitched Instruments**, deviazione standard della quantità di eventi Note On eseguiti da ciascun strumento GeneralMIDI accordato che è usato per riprodurre almeno una nota;
 - **Variability of Number of Independent Voices**, deviazione standard del numero di diversi canali in cui le note sono state eseguite simultaneamente, i resti non sono inclusi in questo calcolo;
 - **Vibrato Prevalence**, numero di note per cui i messaggi di Pitch Bend cambiano direzione almeno due volte diviso per il numero totale di note che hanno messaggi di Pitch Bend associati ad esse;
 - **Violin Fraction**, quantità di eventi Note On appartenenti a *patch* di violino, ossia le *patch* GeneralMIDI 41 o 47;
 - **Voice Separation**, separazione media in semitoni tra i toni medi di canali consecutivi che contengono almeno una nota;
 - **Woodwinds Fraction**, quantità di eventi Note On appartenenti a *patch* di strumenti a fiato, ossia le *patch* General MIDI 41 o 47;

Si costruisce poi la rete con 20 ingressi, due strati da 5 unit nello strato hidden e una unit nello strato di output.

Si impiegano 18 pattern per autore (quindi 36 in totale) nel file di training set e 6 pattern per autore (quindi 12 in totale) nel validation set.

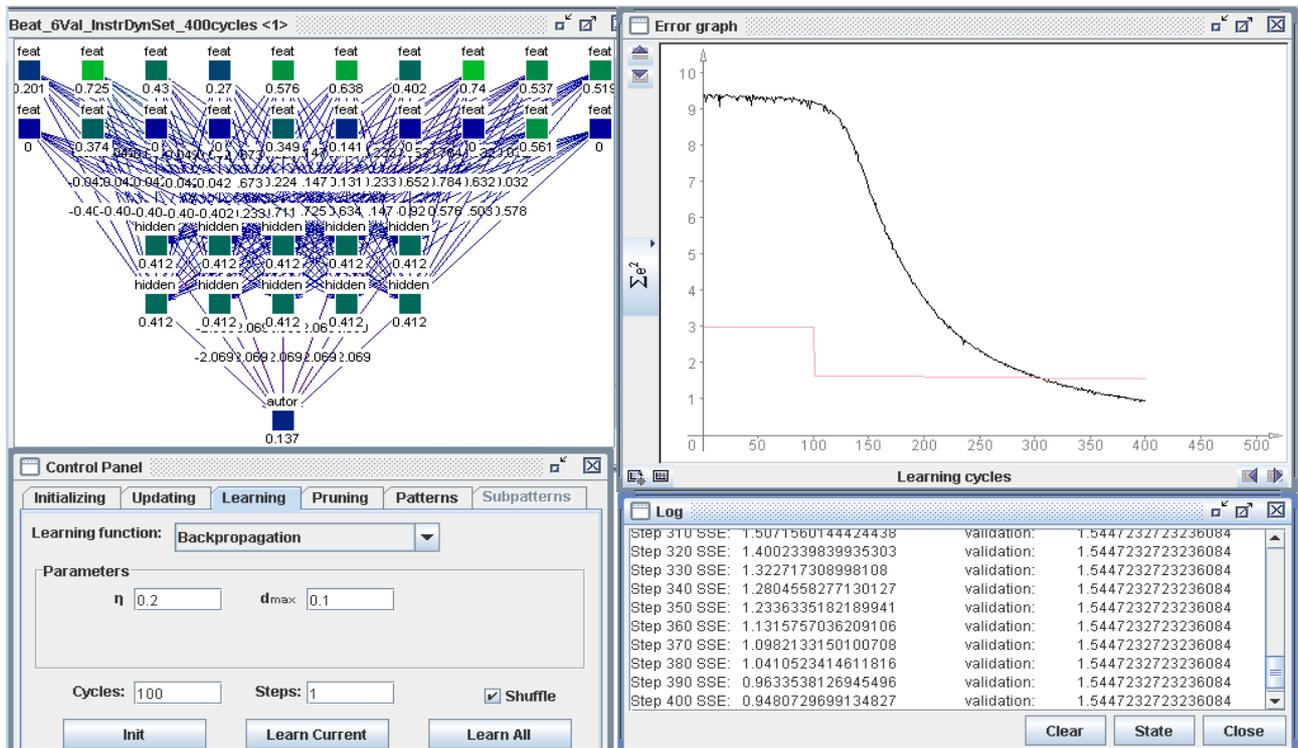


Figura: Beatles vs Led Zeppelin, set di feature strumenti e dinamiche,
 36 pat training 12 pat validation, $\eta=0,2$,
 rete neuronale dopo 400 cicli di apprendimento

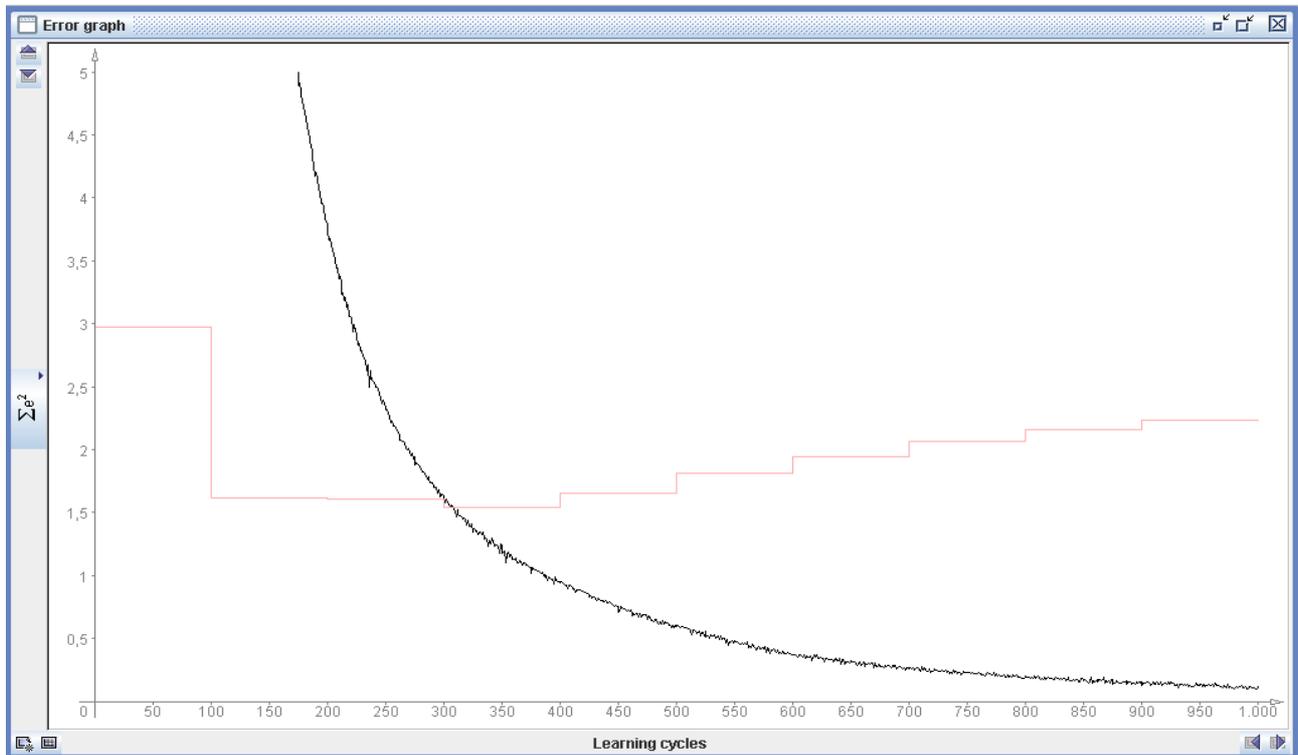


Figura: Beatles vs Led Zeppelin, set di feature strumenti e dinamiche,
 36 pat training 12 pat validation, $\eta=0,2$,
 Error Graph della rete neurale dopo 1000 cicli di apprendimento

Si nota che la linea rossa relativa al validation set raggiunge dopo 400 cicli il minimo valore di $\sum e^2 = 1,5447$, quindi la rete con il miglior livello di generalizzazione è quella ottenuta dopo 400 cicli di apprendimento.

Si modifica poi la distribuzione dei pattern nel training e nel validation set, impiegando 16 pattern per autore (quindi 32 in totale) nel file di training set e 8 pattern per autore (quindi 16 in totale) nel validation set.

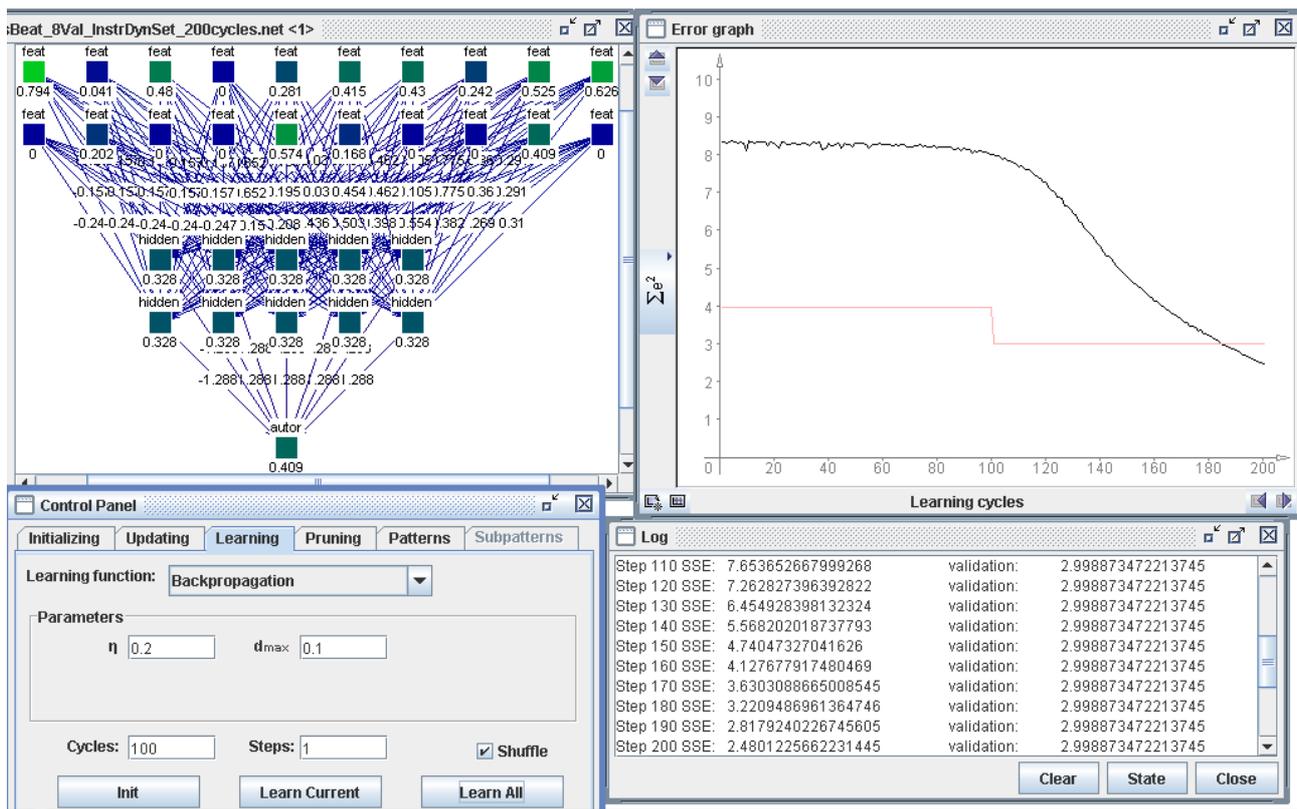


Figura: Beatles vs Led Zeppelin, set di feature strumenti e dinamiche,
 32 pat training 16 pat validation, $\eta=0,2$,
 rete neurale dopo 200 cicli di apprendimento

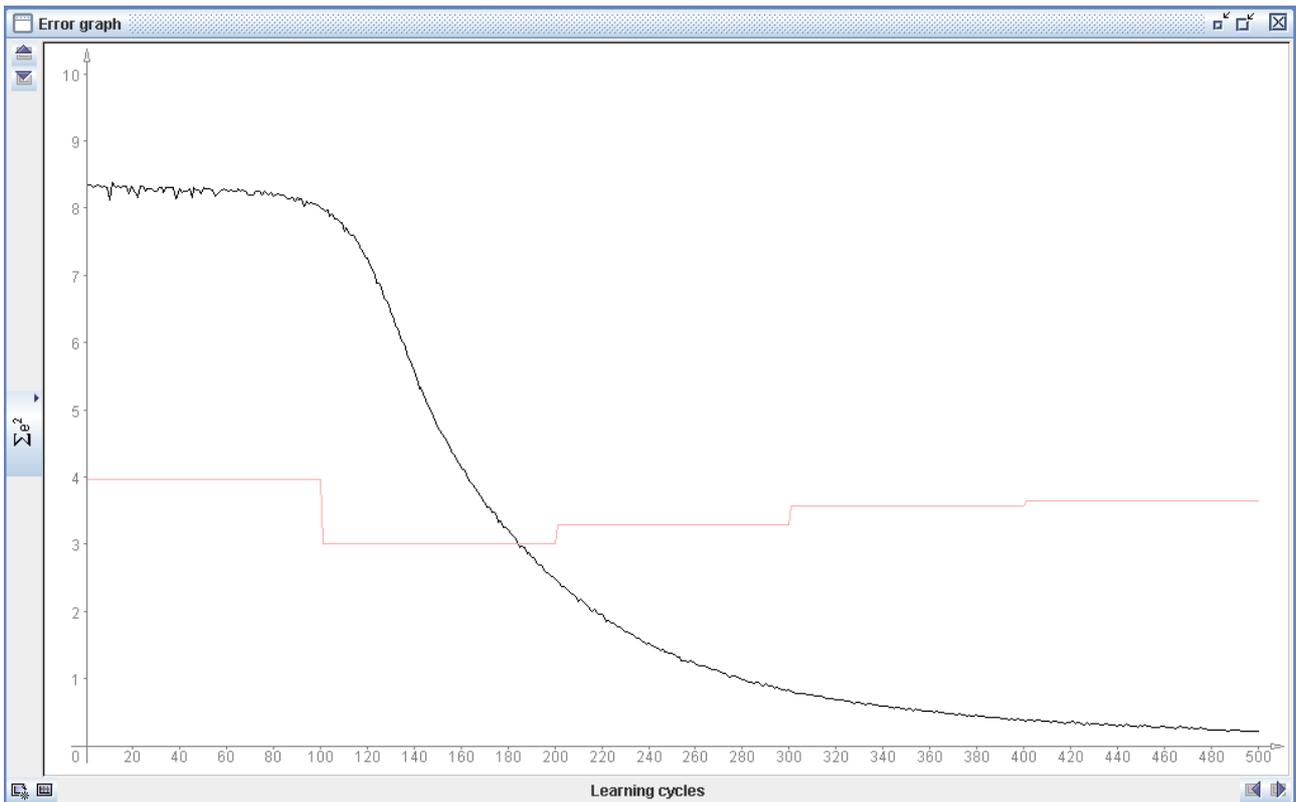


Figura: Beatles vs Led Zeppelin, set di feature strumenti e dinamiche,
 32 pat training 16 pat validation, $\eta=0,2$,
 Error Graph della rete neurale dopo 500 cicli di apprendimento

Si nota che la linea rossa relativa al validation set raggiunge dopo 200 cicli il minimo valore di $\sum e^2 = 2,9988$, quindi la rete con il miglior livello di generalizzazione è quella ottenuta dopo 200 cicli di apprendimento.

In questo caso, a differenza di quanto era avvenuto con il set di feature armonico-melodico, l'aumento della cardinalità del validation set di 4 pattern peggiora considerevolmente il risultato della prova precedente ($\sum e^2 = 1,5447$).

Si selezionano in seguito 20 feature per formare il set di tipo ritmico.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Beatles	Average	Average	Average	Average	Changes	Combin	Compo	Maximum	Note Den	Number	Number	Polyrhyth	Relative	Rhythmic	Rhythmic	Staccato	Strength	Variability	Voice Eq	Variability of
2	Beatles	1,270000	0,770000	2,230000	1,420000	0,000000	1,480000	0,000000	6,350000	4,060000	7,000000	4,000000	0,920000	0,600000	0,000000	0,420000	0,190000	0,750000	1,490000	21,320000	0,420000
3	Beatles	1,020000	0,560000	2,900000	2,350000	0,000000	1,290000	0,000000	3,000000	10,730000	2,990000	5,000000	1,470000	0,600000	0,000000	0,400000	0,720000	0,650000	1,120000	13,080000	0,890000
4	Beatles	0,890000	1,220000	1,490000	3,450000	0,000000	1,160000	0,000000	9,600000	7,430000	2,990000	5,000000	2,210000	1,180000	0,000000	0,320000	0,480000	0,580000	1,920000	11,250000	0,450000
5	Beatles	0,730000	0,460000	3,170000	1,800000	0,000000	1,000000	0,000000	3,130000	9,690000	2,720000	4,000000	1,380000	0,520000	0,000000	0,310000	1,150000	0,530000	1,050000	13,000000	0,960000
6	Beatles	1,350000	0,670000	9,290000	8,200000	0,000000	1,220000	0,000000	6,390000	4,320000	8,000000	4,000000	0,920000	0,610000	0,000000	0,380000	0,160000	0,620000	1,680000	7,870000	0,390000
7	Beatles	1,040000	0,540000	3,230000	2,690000	0,000000	0,960000	0,000000	1,790000	8,110000	4,350000	1,000000	2,450000	0,430000	0,000000	0,230000	0,830000	0,590000	0,930000	17,770000	1,180000
8	Beatles	2,100000	0,810000	2,110000	6,990000	0,000000	1,850000	0,000000	6,920000	3,490000	6,000000	3,000000	1,470000	0,810000	0,000000	0,460000	0,760000	1,060000	2,140000	20,900000	0,490000
9	Beatles	1,110000	0,860000	2,420000	1,210000	0,000000	1,310000	0,000000	1,750000	6,420000	2,720000	4,000000	1,230000	0,470000	0,000000	0,380000	0,720000	0,650000	0,830000	11,670000	0,420000
10	Beatles	1,580000	1,190000	1,470000	3,810000	0,000000	1,550000	0,000000	6,650000	9,010000	2,720000	4,000000	1,470000	2,180000	0,000000	0,410000	0,000000	0,780000	3,230000	10,210000	0,590000
11	Beatles	2,040000	1,200000	1,040000	1,530000	0,000000	1,250000	0,000000	2,660000	4,970000	3,260000	3,000000	2,450000	0,900000	1,230000	0,300000	0,090000	0,670000	1,410000	9,740000	0,540000
12	Beatles	1,090000	1,310000	2,000000	1,040000	1,000000	1,390000	0,000000	3,630000	6,360000	5,000000	3,000000	0,740000	0,480000	2,210000	0,340000	1,640000	0,700000	2,740000	13,000000	0,540000
13	Beatles	0,830000	0,920000	1,230000	3,280000	0,000000	0,720000	0,000000	3,700000	4,900000	2,990000	0,000000	0,740000	1,080000	2,000000	0,210000	2,000000	1,350000	1,850000	8,970000	1,240000
14	Beatles	1,180000	1,190000	1,240000	1,040000	0,000000	0,970000	0,000000	2,030000	5,440000	8,000000	3,000000	0,740000	0,480000	2,940000	0,270000	0,510000	0,480000	1,460000	5,510000	0,380000
15	Beatles	0,420000	0,610000	1,220000	1,840000	0,000000	0,860000	0,000000	2,250000	9,500000	8,000000	1,000000	1,470000	1,170000	2,210000	0,240000	1,290000	0,530000	0,400000	13,000000	0,910000
16	Beatles	1,200000	0,870000	2,090000	1,620000	0,000000	1,660000	0,000000	4,740000	4,240000	8,000000	4,000000	2,760000	1,190000	0,000000	0,450000	0,140000	0,840000	1,480000	7,860000	0,400000
17	Beatles	1,420000	0,890000	2,320000	1,790000	0,000000	1,410000	0,000000	4,560000	3,080000	2,720000	5,000000	0,740000	0,860000	0,000000	0,400000	0,750000	0,710000	1,010000	13,940000	0,400000
18	Beatles	0,810000	0,380000	2,510000	1,470000	0,000000	0,950000	0,000000	1,260000	7,380000	9,000000	3,000000	0,820000	0,660000	0,000000	0,260000	1,430000	0,500000	0,860000	8,770000	1,260000
19	Beatles	1,030000	0,970000	1,700000	3,210000	0,000000	0,950000	0,000000	3,600000	7,470000	2,720000	5,000000	2,210000	0,450000	0,000000	0,280000	1,120000	0,510000	1,130000	14,440000	0,400000
20	Beatles	0,520000	0,520000	1,290000	0,680000	0,000000	1,450000	0,000000	4,050000	8,180000	5,000000	5,000000	1,470000	1,110000	0,000000	0,430000	1,320000	0,750000	0,560000	13,420000	0,390000
21	Beatles	1,120000	0,880000	1,240000	2,510000	1,000000	1,570000	0,000000	9,320000	6,000000	5,000000	5,000000	0,740000	1,270000	0,000000	0,460000	0,660000	0,790000	2,140000	15,390000	0,840000
22	Beatles	1,090000	0,930000	1,980000	3,170000	0,000000	1,950000	0,000000	5,830000	4,860000	7,000000	2,000000	1,580000	0,880000	0,000000	0,450000	0,490000	0,980000	1,000000	14,610000	0,410000
23	Beatles	1,000000	0,500000	3,620000	7,630000	0,000000	0,760000	0,000000	4,140000	3,820000	4,350000	1,000000	2,450000	2,120000	1,230000	0,180000	0,620000	0,460000	1,110000	6,590000	0,650000
24	Beatles	1,490000	0,690000	2,360000	1,690000	0,000000	1,390000	0,000000	3,410000	3,480000	2,990000	4,000000	2,760000	0,760000	0,000000	0,360000	1,980000	0,770000	2,700000	10,960000	0,770000
25	Beatles	1,010000	0,730000	2,070000	3,080000	0,000000	1,420000	0,000000	3,250000	7,100000	5,000000	5,000000	1,470000	0,680000	0,000000	0,440000	1,040000	0,720000	2,050000	13,290000	0,770000
26	LedZeppelin	1,030000	0,480000	2,980000	1,390000	0,000000	0,920000	0,000000	13,320000	8,740000	2,720000	5,000000	1,100000	0,590000	0,000000	0,310000	0,750000	0,460000	2,750000	17,780000	0,900000
27	LedZeppelin	0,890000	1,320000	2,040000	1,200000	0,000000	0,800000	0,000000	4,800000	9,430000	3,530000	3,000000	1,580000	0,570000	0,000000	0,280000	0,160000	0,410000	0,700000	16,890000	0,410000
28	LedZeppelin	0,930000	1,190000	1,370000	3,000000	1,000000	0,760000	0,000000	2,610000	7,600000	2,990000	1,000000	1,380000	0,530000	0,460000	0,240000	1,300000	0,420000	1,220000	15,690000	0,380000
29	LedZeppelin	1,480000	0,860000	2,530000	1,290000	0,000000	1,590000	0,000000	11,960000	6,740000	5,000000	5,000000	1,470000	0,480000	0,000000	0,460000	0,290000	0,810000	1,150000	14,660000	1,090000
30	LedZeppelin	1,050000	0,380000	1,020000	4,260000	1,000000	0,690000	1,000000	4,950000	3,560000	3,260000	0,000000	1,230000	0,720000	1,170000	0,210000	1,440000	1,300000	2,030000	25,130000	0,470000
31	LedZeppelin	1,540000	1,260000	1,950000	4,040000	1,000000	0,920000	1,000000	6,370000	7,770000	4,350000	2,000000	2,210000	0,400000	0,000000	0,250000	0,540000	0,530000	2,180000	9,210000	0,460000
32	LedZeppelin	1,010000	0,540000	1,840000	2,460000	0,000000	1,060000	0,000000	4,880000	4,480000	7,000000	4,000000	1,580000	0,450000	0,000000	0,340000	0,820000	0,550000	1,330000	18,120000	1,350000
33	LedZeppelin	0,980000	0,590000	1,600000	1,770000	1,000000	1,120000	0,000000	1,430000	8,840000	7,000000	7,000000	1,580000	1,140000	0,000000	0,370000	0,440000	0,580000	0,660000	17,170000	1,100000
34	LedZeppelin	0,800000	0,910000	1,870000	1,490000	1,000000	0,670000	0,000000	3,080000	6,400000	3,260000	1,000000	0,950000	0,730000	2,630000	0,200000	0,690000	0,390000	0,680000	10,490000	1,260000
35	LedZeppelin	0,900000	0,600000	2,390000	3,890000	1,000000	0,900000	0,000000	5,000000	4,640000	2,720000	3,000000	0,920000	0,780000	0,000000	0,300000	1,420000	0,450000	1,080000	13,330000	0,580000
36	LedZeppelin	1,220000	0,820000	1,630000	2,850000	0,000000	1,390000	0,000000	1,750000	5,050000	2,720000	5,000000	2,210000	0,570000	0,000000	0,430000	0,000000	0,710000	0,620000	26,930000	0,830000
37	LedZeppelin	0,570000	0,590000	0,940000	1,230000	0,000000	0,640000	0,000000	4,680000	7,940000	2,720000	0,000000	1,230000	1,230000	3,070000	0,200000	2,020000	1,230000	1,390000	12,970000	1,000000
38	LedZeppelin	0,900000	0,790000	1,950000	1,720000	1,000000	1,140000	0,000000	4,210000	3,660000	8,000000	4,000000	0,920000	0,570000	0,000000	0,360000	1,420000	0,580000	1,780000	16,400000	0,990000
39	LedZeppelin	1,230000	1,480000	17,210000	9,390000	1,000000	1,850000	0,000000	21,240000	7,560000	4,000000	4,000000	1,840000	1,440000	0,000000	0,490000	0,860000	1,020000	3,640000	8,680000	1,850000
40	LedZeppelin	0,720000	1,140000	1,180000	2,260000	0,000000	0,970000	0,000000	4,950000	5,340000	2,720000	2,000000	0,740000	0,960000	1,000000	0,270000	1,570000	0,530000	1,470000	10,180000	0,470000
41	LedZeppelin	0,450000	0,440000	2,300000	6,010000	1,000000	0,700000	0,000000	1,880000	3,440000	7,070000	6,00									

- frequenze dei due *beat bin* dei picchi con la frequenza più alta;
- **Compound Or Simple Meter**, settare a 1 se il metro iniziale è composto, ossia con il numeratore pari a 3 o ad un suo multiplo, o settare a 0 se il metro iniziale è semplice;
 - **Maximum Note Duration**, durata in secondi della nota più lunga;
 - **Note Density**, numero medio di note per secondo;
 - **Number of Moderate Pulses**, numero di picchi di beat con frequenza normalizzata superiore a 0,01;
 - **Number of Strong Pulses**, numero di picchi di beat con frequenza normalizzata superiore a 0,1;
 - **Polyrhythms**, numero di picchi di beat con frequenze almeno al 30% della frequenza più alta di cui i valori dei *bin* non sono interi multipli o fattori (usando solo 1, 2, 3, 4, 5, 6, e 8 con un errore accettabile di +/- 3 *bin*) del valore del *bin* relativo al picco di beat con la frequenza più alta (questo numero è poi diviso per il numero totale di *beat bin* con frequenze maggiori del 30% della frequenza più alta);
 - **Relative Note Density of Highest Line**, numero di eventi Note On nel canale con la media di tono più alta divisa per il numero medio di eventi Note On in tutti i canali che contengono almeno una nota;
 - **Rhythmic Looseness**, ampiezza media (in beat per minuto) dei picchi dell'istogramma dei beat (l'ampiezza è misurata per tutti i picchi con frequenze alte almeno il 30% del picco più alto ed è definita dalla distanza tra i punti sui picchi in questione che sono il 30% dell'altezza del picco più alto);
 - **Rhythmic Variability**, deviazione standard dei valori dei *bin* eccetto i i primi 40 vuoti;
 - **Staccato Incidence**, numero di note con durata minore di un decimo di secondo diviso per il numero totale di note nel pezzo;
 - **Strength of Strongest Rhythmic Pulse**, frequenza del *beat bin* con la

frequenza più alta;

- **Variability of Note Duration**, deviazione standard della durata delle note in secondi;
- **Voice Equality - Note Duration**, deviazione standard della durata totale delle note in secondi in ogni canale che contiene almeno una nota;
- **Variability of Time Between Attacks**, deviazione standard dei tempi in secondi tra eventi Note On a prescindere dal canale.

Si costruisce poi la rete con 20 ingressi, due strati di 5 unit nello strato hidden e una unit nello strato di output.

Si impiegano 18 pattern per autore (quindi 36 in totale) nel file di training set e 6 pattern per autore (quindi 12 in totale) nel validation set.

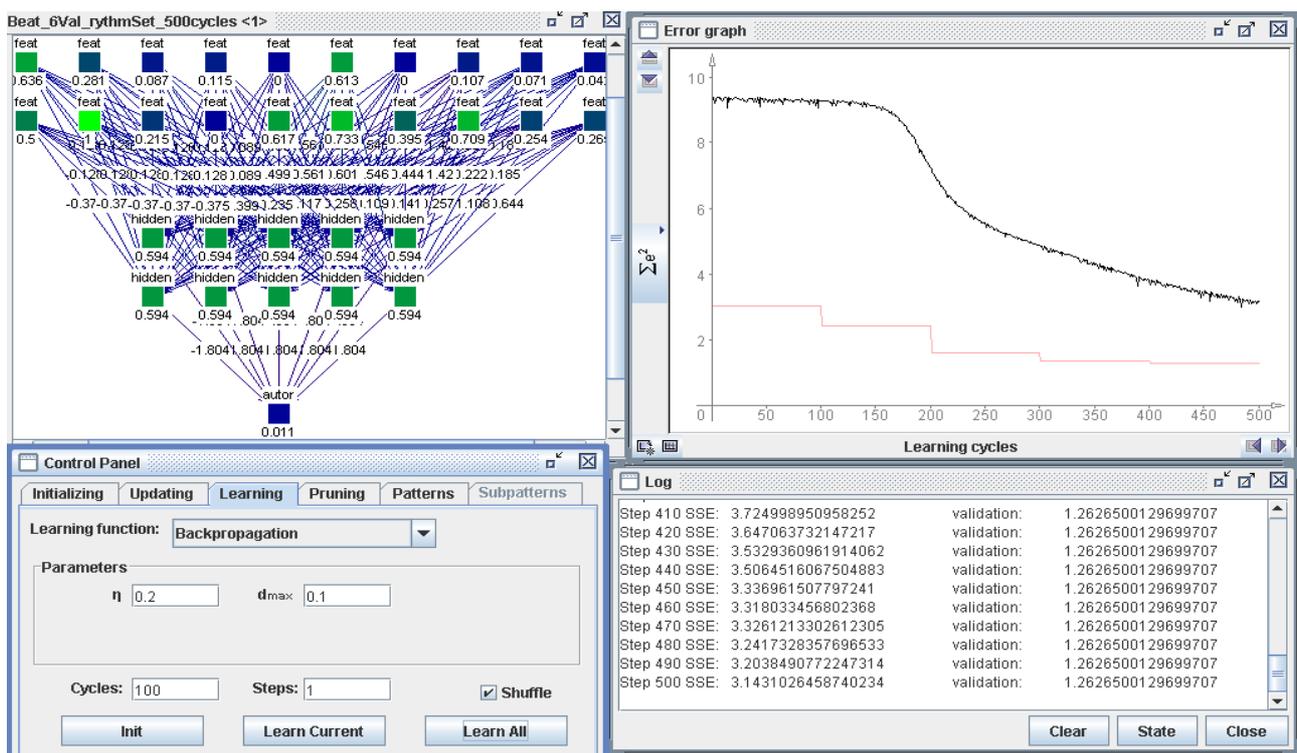


Figura: Beatles vs Led Zeppelin, set di feature ritmico,
36 pat training 12 pat validation, $\eta=0,2$,
rete neuronale dopo 500 cicli di apprendimento

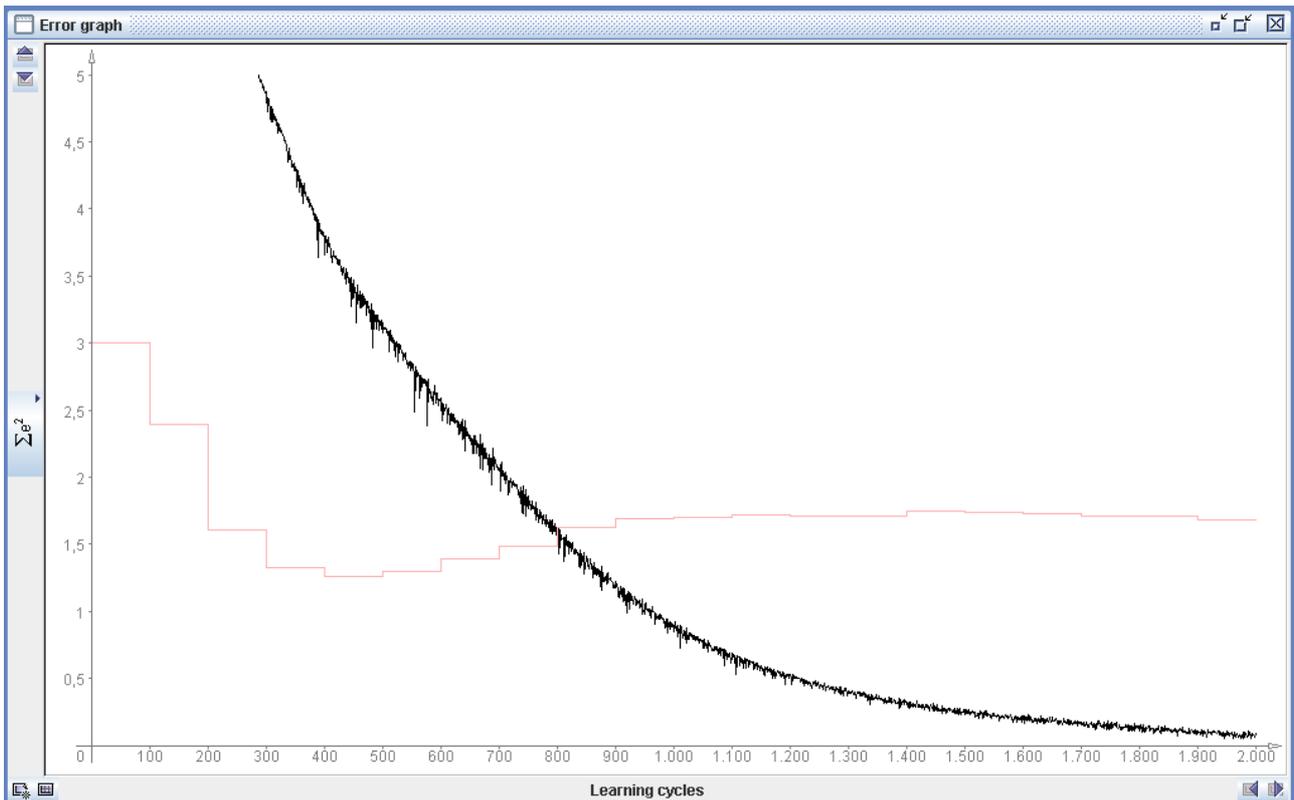


Figura: Beatles vs Led Zeppelin, set di feature ritmico,
 36 pat training 12 pat validation, $\eta=0,2$,
 Error Graph della rete neuronale dopo 2000 cicli di apprendimento

Si nota che la linea rossa relativa al validation set raggiunge dopo 500 cicli il minimo valore di $\sum e^2 = 1,2626$, quindi la rete con il miglior livello di generalizzazione è quella ottenuta dopo 500 cicli di apprendimento.

Si modifica poi la distribuzione dei pattern nel training e nel validation set, impiegando 16 pattern per autore (quindi 32 in totale) nel file di training set e 8 pattern per autore (quindi 16 in totale) nel validation set.

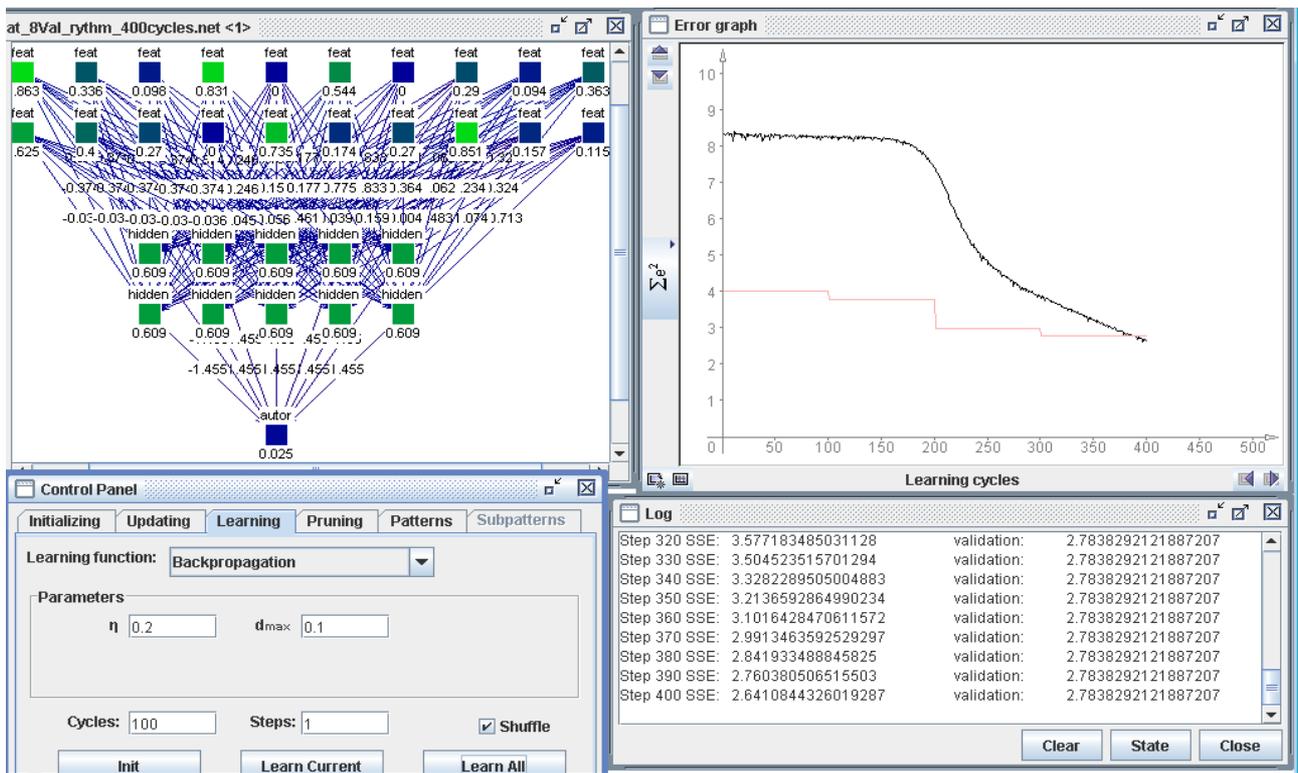


Figura: Beatles vs Led Zepelin, set di feature ritmico,
 32 pat training 16 pat validation, $\eta=0,2$,
 rete neuronale dopo 400 cicli di apprendimento

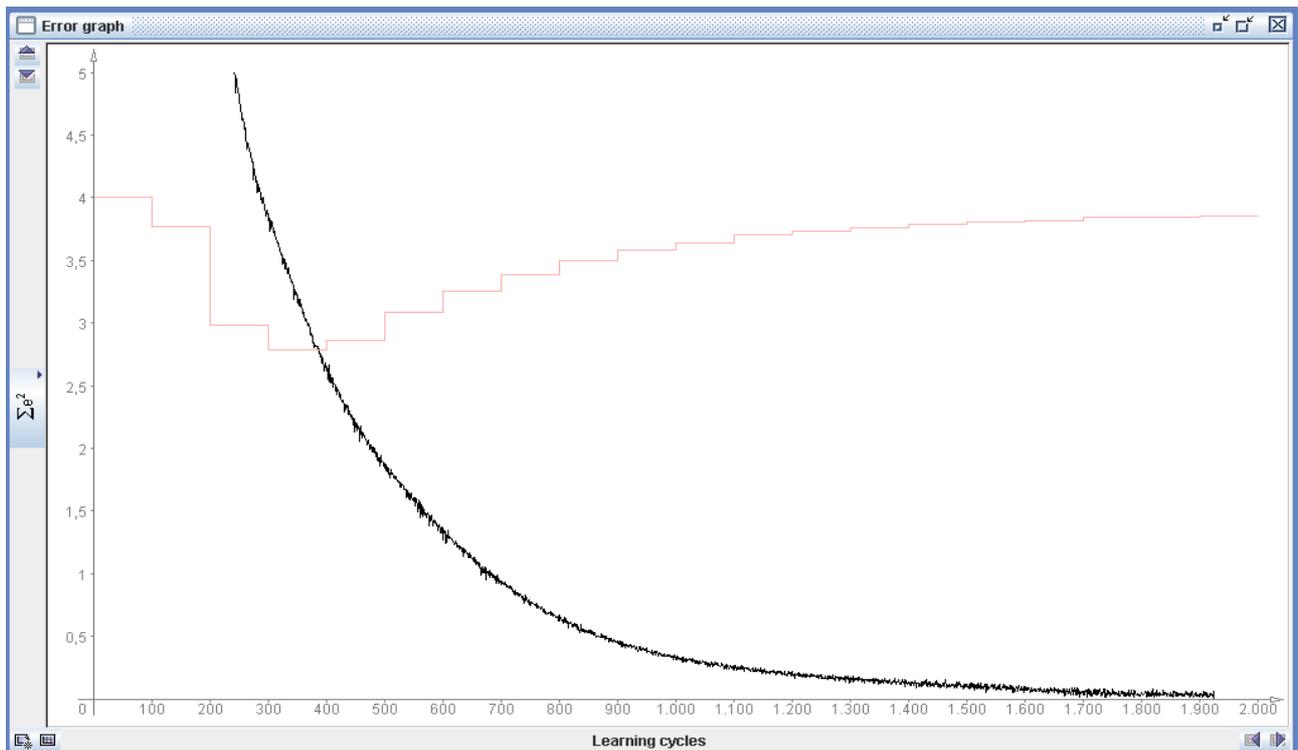


Figura: Beatles vs Led Zeppelin, set di feature ritmico,

32 pat training 16 pat validation, $\eta=0,2$,

Error Graph della rete neurale dopo 2000 cicli di apprendimento

Si nota che la linea rossa relativa al validation set raggiunge dopo 400 cicli il minimo valore di $\sum e^2 = 2,7838$, quindi la rete con il miglior livello di generalizzazione è quella ottenuta dopo 400 cicli di apprendimento.

A questo punto si decide di costruire un ultimo **set misto**. Si analizzano le tre tabelle dei **tre gruppi** da 20 di feature descritti e **per ognuno** di essi si scelgono le **4 feature** giudicate più significative.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Beatles	0.000000	1.910000	2.470000	1.370000	0.000000	9.790000	1.890000	0.790000	21.320000	0.000000	1.480000	0.420000
2	Beatles	0.000000	8.160000	1.170000	1.460000	0.000000	12.500000	1.900000	0.800000	13.080000	0.000000	1.290000	0.890000
3	Beatles	0.000000	9.290000	2.080000	1.300000	0.000000	11.420000	2.260000	0.410000	11.250000	0.000000	1.160000	0.450000
4	Beatles	0.000000	9.870000	0.000000	2.360000	2.890000	10.600000	0.840000	0.650000	13.000000	0.000000	1.000000	0.960000
5	Beatles	0.000000	2.400000	1.160000	1.060000	0.000000	11.690000	1.580000	0.880000	7.870000	0.000000	1.220000	0.390000
6	Beatles	0.000000	3.490000	0.980000	2.700000	0.000000	9.240000	2.480000	1.220000	17.770000	0.000000	0.960000	1.180000
7	Beatles	0.000000	1.980000	1.000000	2.230000	0.000000	10.600000	1.680000	1.240000	20.900000	0.000000	1.850000	0.490000
8	Beatles	0.000000	2.720000	0.490000	2.110000	0.000000	7.880000	1.600000	0.620000	11.670000	0.000000	1.310000	0.420000
9	Beatles	0.000000	1.080000	0.440000	0.940000	9.670000	9.790000	1.720000	1.110000	10.210000	0.000000	1.550000	0.590000
10	Beatles	0.000000	1.930000	0.400000	2.680000	0.000000	11.960000	2.490000	0.920000	9.740000	0.000000	1.250000	0.540000
11	Beatles	0.000000	9.560000	0.790000	2.320000	0.000000	12.500000	1.760000	0.700000	13.000000	1.000000	1.390000	0.540000
12	Beatles	0.000000	9.420000	1.410000	2.410000	12.740000	9.790000	1.870000	0.890000	8.970000	0.000000	0.720000	1.240000
13	Beatles	0.000000	3.830000	0.840000	2.740000	0.000000	7.610000	1.530000	0.480000	5.510000	0.000000	0.970000	0.380000
14	Beatles	0.000000	4.480000	2.480000	2.610000	0.000000	7.610000	1.780000	0.580000	13.000000	0.000000	0.860000	0.910000
15	Beatles	0.000000	2.670000	1.330000	1.800000	0.000000	11.960000	1.950000	0.520000	7.860000	0.000000	1.660000	0.400000
16	Beatles	0.000000	0.860000	1.840000	1.900000	0.000000	9.240000	1.990000	0.860000	13.940000	0.000000	1.410000	0.400000
17	Beatles	0.000000	5.660000	2.000000	3.040000	0.000000	8.150000	1.000000	0.790000	8.770000	0.000000	0.950000	1.260000
18	Beatles	0.000000	5.800000	2.060000	1.670000	0.000000	10.870000	2.420000	0.470000	14.440000	0.000000	0.950000	0.400000
19	Beatles	0.000000	2.910000	1.550000	2.030000	0.000000	8.700000	2.270000	0.940000	13.420000	0.000000	1.450000	0.390000
20	Beatles	0.000000	1.150000	0.460000	2.830000	1.520000	10.870000	0.860000	0.810000	15.390000	1.000000	1.570000	0.840000
21	Beatles	0.000000	1.510000	0.490000	1.530000	0.000000	10.330000	1.840000	0.750000	14.610000	0.000000	1.950000	0.410000
22	Beatles	0.000000	5.890000	0.000000	2.340000	0.000000	10.600000	2.130000	1.140000	6.590000	0.000000	0.760000	0.650000
23	Beatles	0.000000	3.460000	0.000000	2.220000	0.000000	9.240000	1.420000	0.790000	10.960000	0.000000	1.390000	0.770000
24	Beatles	0.000000	7.160000	1.660000	1.960000	0.000000	11.960000	1.650000	0.700000	13.290000	0.000000	1.420000	0.770000
25	LedZeppelin	0.000000	3.070000	2.090000	1.710000	0.000000	15.220000	0.940000	0.620000	17.780000	0.000000	0.920000	0.900000
26	LedZeppelin	0.000000	2.820000	2.600000	1.160000	6.280000	13.590000	1.020000	0.410000	16.890000	0.000000	0.800000	0.410000
27	LedZeppelin	0.350000	0.700000	2.030000	1.590000	13.490000	11.140000	1.490000	0.760000	15.690000	1.000000	0.760000	0.380000
28	LedZeppelin	0.000000	0.000000	2.040000	1.530000	0.000000	11.690000	0.880000	1.130000	14.660000	0.000000	1.590000	1.090000
29	LedZeppelin	0.000000	3.400000	1.840000	1.120000	14.220000	12.780000	1.660000	1.320000	25.130000	1.000000	0.690000	0.470000
30	LedZeppelin	0.000000	0.480000	2.050000	2.270000	0.000000	14.140000	1.540000	0.540000	9.210000	1.000000	0.920000	0.460000
31	LedZeppelin	0.000000	1.340000	2.240000	2.450000	16.550000	8.700000	1.840000	0.690000	18.120000	0.000000	1.060000	1.350000
32	LedZeppelin	0.000000	1.090000	0.830000	1.660000	0.000000	10.870000	1.250000	0.800000	17.170000	1.000000	1.120000	1.100000
33	LedZeppelin	0.650000	4.120000	1.850000	1.700000	13.800000	14.140000	2.060000	0.760000	10.490000	1.000000	0.670000	1.260000
34	LedZeppelin	0.000000	1.500000	1.960000	1.580000	10.800000	14.950000	1.970000	1.150000	13.330000	1.000000	0.900000	0.580000
35	LedZeppelin	0.000000	1.160000	1.950000	1.980000	0.000000	9.240000	1.720000	0.650000	26.930000	0.000000	1.390000	0.830000
36	LedZeppelin	0.000000	4.680000	2.020000	1.370000	2.820000	11.960000	1.930000	0.480000	12.970000	0.000000	0.640000	1.000000
37	LedZeppelin	0.190000	2.610000	2.040000	1.720000	12.550000	12.230000	1.470000	1.020000	16.400000	1.000000	1.140000	0.990000
38	LedZeppelin	0.000000	3.250000	1.040000	1.260000	12.000000	14.410000	1.880000	1.640000	8.680000	1.000000	1.850000	1.850000
39	LedZeppelin	0.280000	3.630000	1.910000	2.060000	13.030000	11.690000	1.980000	1.330000	10.180000	0.000000	0.970000	0.470000
40	LedZeppelin	0.000000	2.400000	1.010000	1.420000	4.600000	10.060000	1.290000	1.730000	9.930000	1.000000	0.700000	0.400000
41	LedZeppelin	0.000000	1.240000	0.570000	1.740000	0.000000	15.220000	1.670000	1.140000	8.880000	0.000000	1.290000	0.550000
42	LedZeppelin	0.000000	0.410000	1.080000	2.200000	0.000000	9.790000	0.950000	0.700000	20.330000	0.000000	0.910000	0.650000
43	LedZeppelin	0.000000	0.780000	1.670000	0.820000	8.430000	14.410000	1.840000	0.750000	14.900000	1.000000	1.210000	0.810000
44	LedZeppelin	0.290000	3.590000	2.200000	2.110000	12.150000	12.500000	1.850000	1.150000	10.820000	0.000000	0.690000	0.650000
45	LedZeppelin	0.360000	3.500000	1.970000	0.790000	12.820000	8.150000	1.850000	0.870000	14.710000	0.000000	0.500000	0.690000
46	LedZeppelin	0.290000	2.990000	0.800000	2.490000	12.380000	14.680000	1.880000	0.560000	25.990000	0.000000	0.540000	1.140000
47	LedZeppelin	0.000000	4.380000	1.630000	0.870000	2.090000	17.130000	1.580000	1.130000	20.620000	1.000000	0.990000	1.050000
48	LedZeppelin	0.000000	0.550000	0.800000	2.400000	0.000000	13.050000	1.750000	0.910000	26.860000	1.000000	1.300000	0.640000

Figura: Beatles vs Led Zeppelin, tabella contenente

il set di feature misto

Ecco di seguito le 12 feature scelte per formare il set misto.

Le prime quattro appartengono al set riguardante le dinamiche e gli strumenti utilizzati:

- **Vibrato Prevalence**, numero di note per cui i messaggi di Pitch Bend cambiano direzione almeno due volte diviso per il numero totale di note che hanno messaggi di Pitch Bend associati ad esse;
- **Average Note To Note Dynamics Change**, differenza media di intensità, calcolata in MIDI velocity unit, tra una nota e la seguente nel medesimo canale;
- **Electric Instrument Fraction**, quantità di eventi Note On appartenenti a *patch* di strumenti elettrici, ossia le *patch* GeneralMIDI 5, 6, 17, 19, dalla 27 alla 32 o dalla 34 alla 40;
- **Importance of Middle Register**, quantità di eventi Note On con toni MIDI dallo 55 al 72.

Le seconde quattro appartengono al set armonico-melodico:

- **Average Range of Glissandos**, *range* medio di Pitch Bend, dove per *range* si intende il più alto valore ottenuto dalla differenza assoluta tra 64 ed il secondo data byte di tutti gli eventi MIDI di Pitch Bend che cadono tra l'evento di Note On e l'evento di Note Off di ogni nota;
- **Pitch Variety**, numero di note usate almeno una volta;
- **Direction of Motion**, quantità di intervalli melodici ascendenti piuttosto che discendenti;
- **Stepwise Motion**, quantità di intervalli melodici che corrispondono ad una seconda minore o maggiore.

Le ultime quattro appartengono al set ritmico:

- **Voice Equality - Note Duration**, deviazione standard della durata totale delle note in secondi in ogni canale che contiene almeno una nota;
- **Changes of Meter**, settare a 1 se il metro ritmico è cambiato una o più volte

durante il pezzo;

- **Combined Strength of Two Strongest Rhythmic Pulses**, somma delle frequenze dei due *beat bin* dei picchi con la frequenza più alta;
- **Variability of Time Between Attacks**, deviazione standard dei tempi in secondi tra eventi Note On a prescindere dal canale.

Si costruisce poi la rete con 12 ingressi, due strati da 3 unit nello strato hidden e una unit nello strato di output.

Si impiegano 18 pattern per autore (quindi 36 in totale) nel file di training set e 6 pattern per autore (quindi 12 in totale) nel validation set.

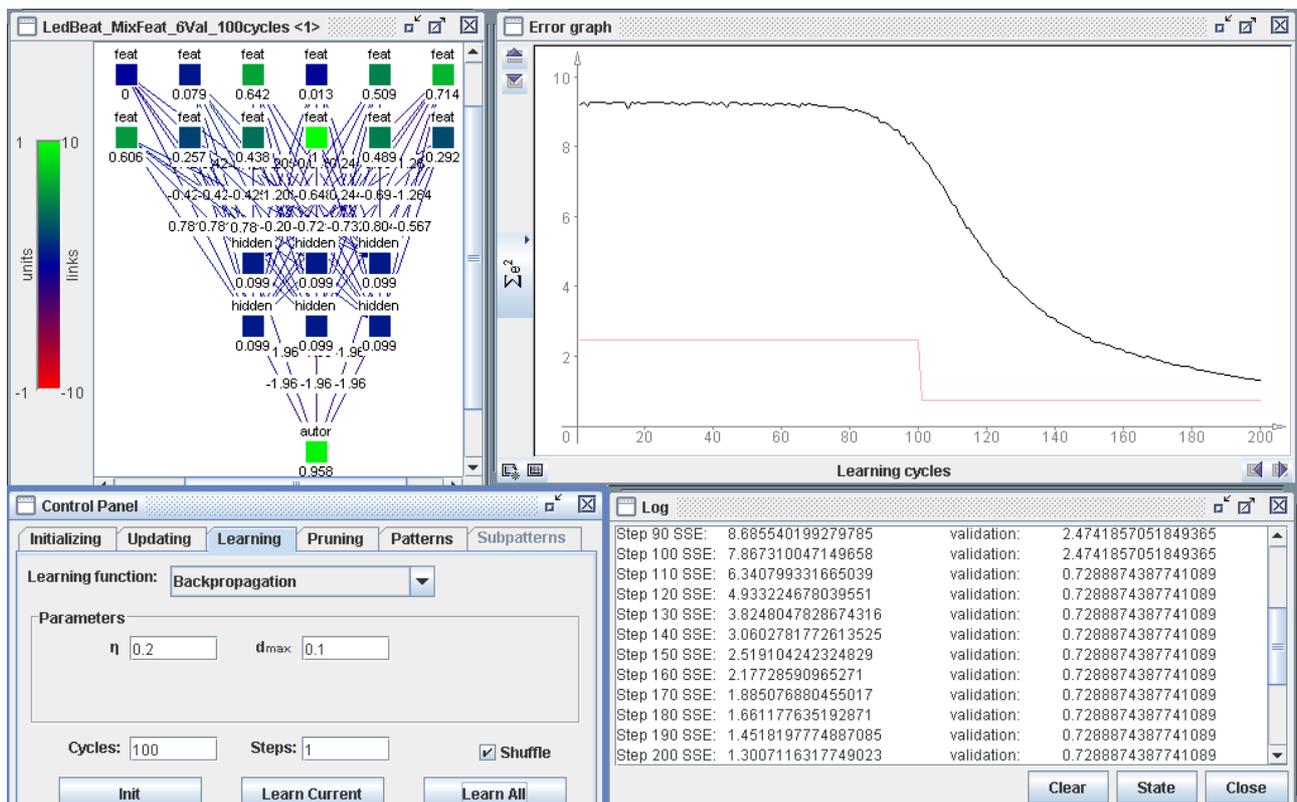


Figura: Beatles vs Led Zepplin, set di feature misto,
36 pat training 12 pat validation, $\eta=0,2$,
rete neuronale dopo 200 cicli di apprendimento

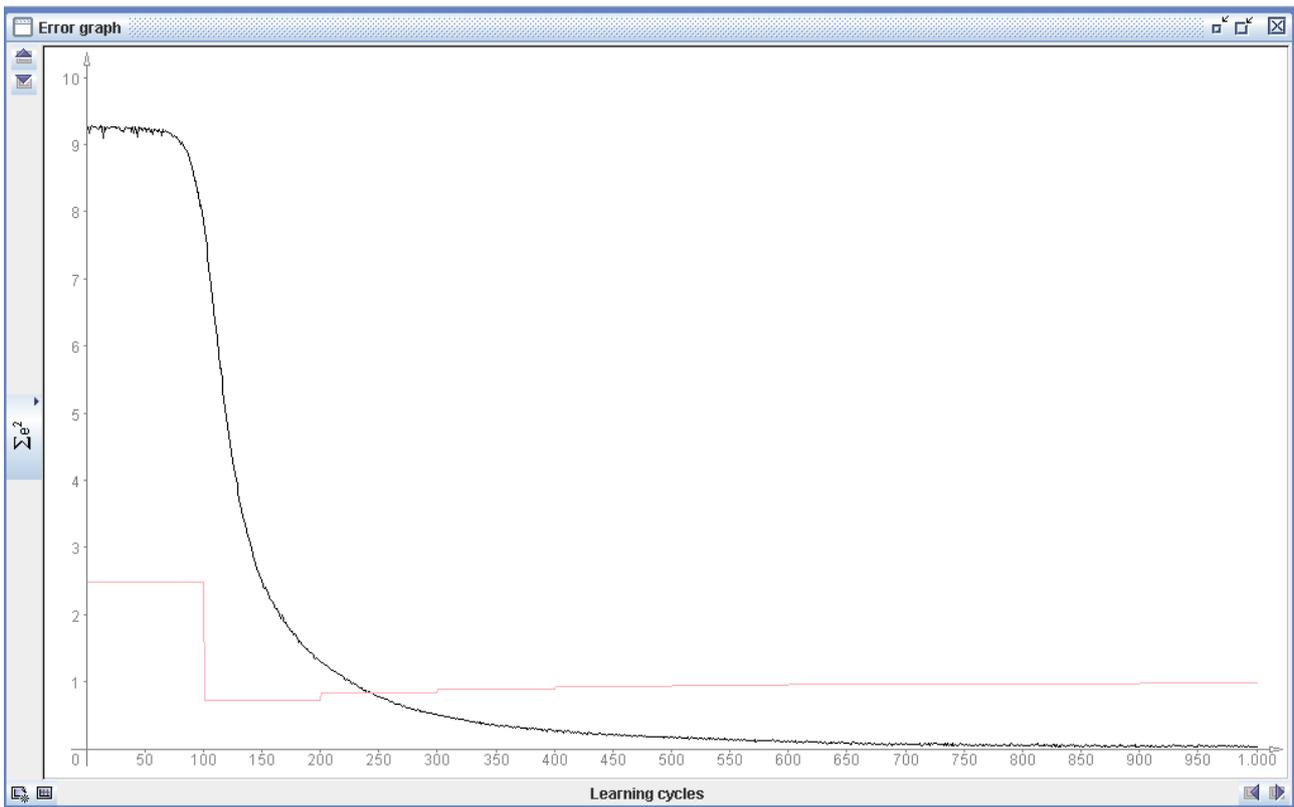


Figura: Beatles vs Led Zeppelin, set di feature misto,
 36 pat training 12 pat validation, $\eta=0,2$,
 rete neurale dopo 1000 cicli di apprendimento

Si nota che la linea rossa relativa al validation set raggiunge dopo 500 cicli il minimo valore di $\sum e^2 = 0,7288$, quindi la rete con il miglior livello di generalizzazione è quella ottenuta dopo 500 cicli di apprendimento.

Si modifica poi la distribuzione dei pattern nel training e nel validation set, impiegando 16 pattern per autore (quindi 32 in totale) nel file di training set e 8 pattern per autore (quindi 16 in totale) nel validation set.

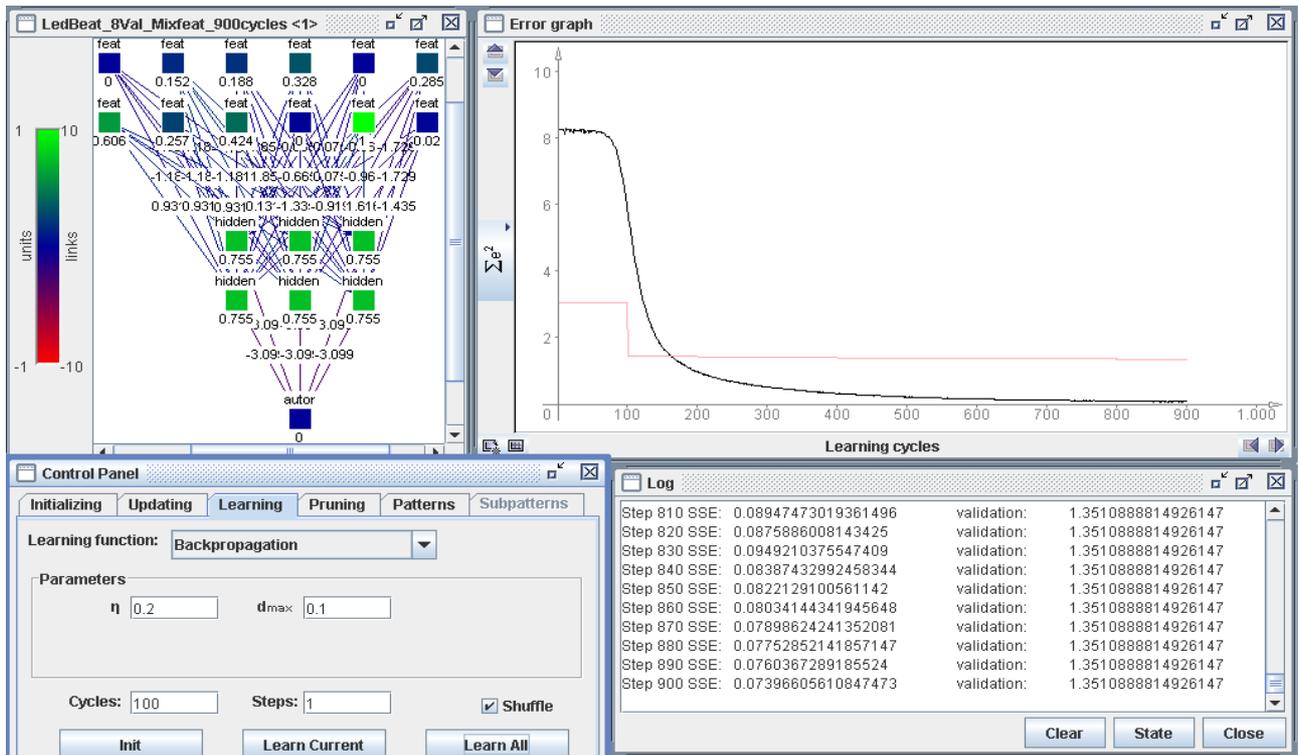


Figura: Beatles vs Led Zeppelin, set di feature ritmico,
 32 pat training 16 pat validation, $\eta=0,2$,
 rete neuronale dopo 900 cicli di apprendimento

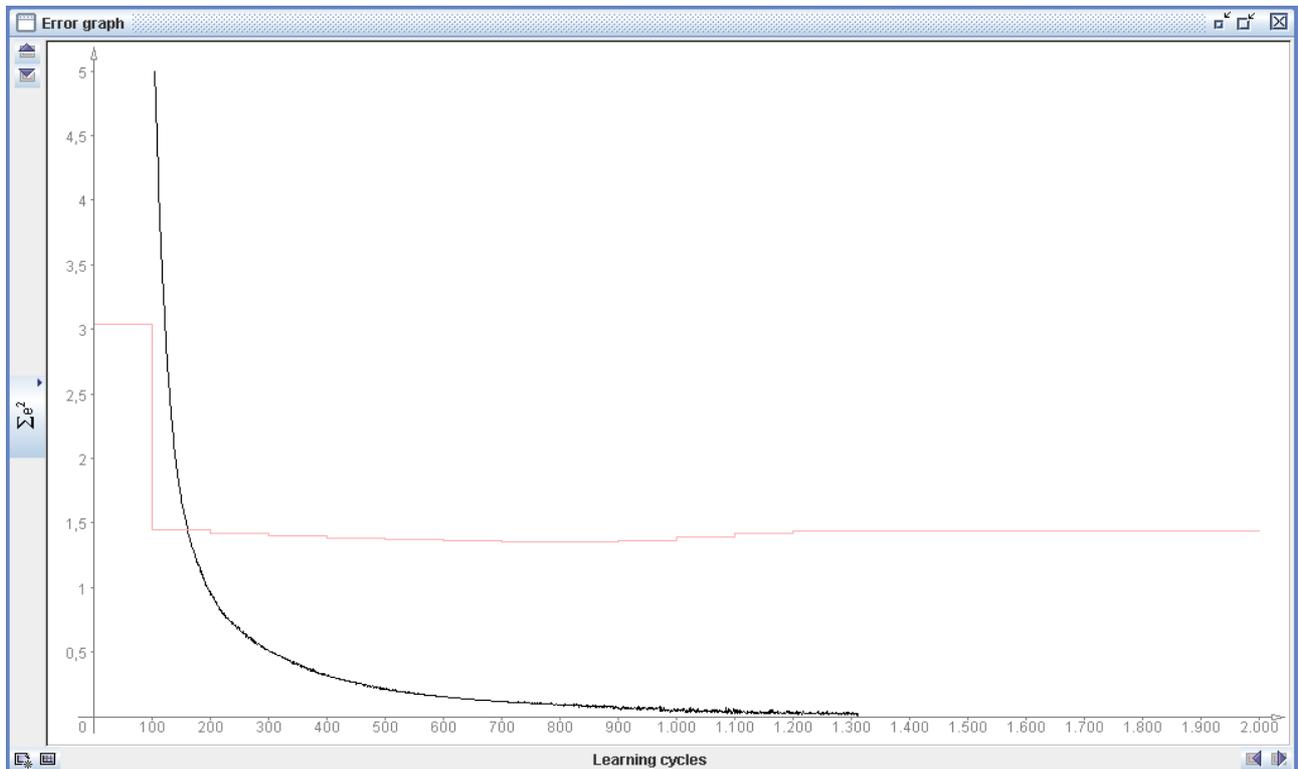


Figura: Beatles vs Led Zeppelin, set di feature ritmico,

32 pat training 16 pat validation, $\eta=0,2$,

Error Graph della rete neuronale dopo 2000 cicli di apprendimento

Si nota che la linea rossa relativa al validation set raggiunge dopo 900 cicli il minimo valore di $\sum e^2 = 1,3510$, quindi la rete con il miglior livello di generalizzazione è quella ottenuta dopo 900 cicli di apprendimento.

Riassunto Risultati con i 4 Differenti Set di Feature

	Set mel-arm	Set strum e din	Set ritmico	Set misto
36 pat training 12 pat validation	$\sum e^2 = 1,7030$	$\sum e^2 = 1,5447$	$\sum e^2 = 1,2626$	$\sum e^2 = 0,7288$
32 pat training 16 pat validation	$\sum e^2 = 1,7315$	$\sum e^2 = 2,9988$	$\sum e^2 = 2,7838$	$\sum e^2 = 1,3510$

E' evidente che il set misto è il più efficiente anche se è dotato di ben 8 feature in meno rispetto agli altri tre.

Confrontando questi risultati con quelli del paragrafo precedente si scopre che, **come ipotizzato, il riconoscimento dei Beatles accostati ai Led Zeppelin ha raggiunto livelli di efficienza più alta rispetto al riconoscimento degli stessi accostati ai Rolling Stones.**

Con una distribuzione di 36 pattern nel training set e 12 nel validation set il miglior risultato negli esperimenti del paragrafo precedente è $\sum e^2 = 2,4309$ con la modifica del parametro η portato a 0,1.

Invece con una distribuzione di 32 pattern nel training set e 16 nel validation set il miglior risultato negli esperimenti del paragrafo precedente è $\sum e^2 = 3,3050$.

Conclusioni e Sviluppi Futuri

Considerando gli esperimenti svolti si può concludere che in linea generale i risultati ottenuti sono buoni. L'andamento delle linee relative al training e al validation set presenti nei grafici rispetta, con alcune eccezioni, la configurazione del grafico standard che si ottiene con l'applicazione dell'algoritmo di Backpropagation ad una rete Feedforward. Le reti raggiungono discreti livelli di apprendimento anche se esistono significative differenze a seconda di quale coppia di compositori si considera. Le prove relative a Bach e Mozart, nonostante i pattern disponibili per la rete siano in numero nettamente inferiore rispetto agli esperimenti successivi, sono quelli che portano ai risultati migliori. Questo suggerisce che si sono trovate feature che catturano aspetti distintivi dei due compositori. In ordine di efficienza seguono le reti addestrate per i confronti diretti tra Luis Armstrong e Chet Baker e tra Beatles e Led Zeppelin con il set misto di feature. Seguono poi tutti gli altri. L'esperienza dal risultato più fallimentare riguarda il confronto tra Chet Baker e l'insieme di altri trombettisti, in cui la linea relativa al validation set non raggiunge nessun minimo se non si interviene sul parametro η .

Alla luce di questi risultati sostanzialmente soddisfacenti si può pensare a nuovi futuri esperimenti in cui aumentare il livello di complessità dei compiti da far svolgere alle reti neurali costruite ai fini della classificazione.

Ad esempio si potrebbe pensare di aumentare progressivamente il numero di autori da analizzare per ogni prova, incrementando di conseguenza anche il numero di unit nello strato di output.

Un altro interessante studio consisterebbe nel coniugare feature di alto livello estratte con JSymbolic con feature culturali o di basso livello estratte con altri programmi

appositi, alcuni disponibili anche nel pacchetto software JMIR. Le feature di basso livello, ad esempio, potrebbero essere impiegate per riconoscere la voce dei cantanti usando file "wave format".

Guardando, infine, molto avanti nel processo di ricerca si possono immaginare sistemi intelligenti che coniugano il processo di classificazione con quello di composizione. Essi infatti riconoscendo l'autore di certe opere mai analizzate potrebbero, sulla base delle feature distintive possedute, creare nuovi pattern musicali o addirittura nuovi brani caratterizzati dallo stile dell'autore in questione. Questo processo incrementerebbe di interesse se potesse avvenire in real-time permettendo una interazione utente-macchina di tipo musicale. Utilizzando altre tecnologie sistemi di questo tipo sono già stati costruiti. In tale direzione infatti stanno andando gli studi di molti ricercatori : George Tzanetakis, Francois Pachet, Georg Essl, Perry Cook, Douk Eck, Judy A. Franklin, P.M. Todd, Tim Thomson, Dominic Hornel, Wolfram Menzel, e tanti altri; alcuni risultati sono sorprendenti anche se la ricerca in questo campo ha ancora molti passi da fare.

Bibliografia e Sitografia

1.0 Reti Neuronal;

[001] [“Neural Networks”](#), Andrea Roli

1.1 Reti Neuronal Biologiche;

[002] http://it.wikipedia.org/wiki/Rete_neurale

[003] http://www.irccsdebellis.it/html/reti_neurali/teoria_delle_reti_neurali_1.htm

[004] http://it.wikipedia.org/wiki/Neurone#Trasmissione_dell.27impulso_elettrico

[005] <http://it.wikipedia.org/wiki/Neurone#Anatomia>

[006] <http://it.wikipedia.org/wiki/Assone>

[007] [http://it.wikipedia.org/wiki/Dendrite_\(biologia\)](http://it.wikipedia.org/wiki/Dendrite_(biologia))

[008] http://it.wikipedia.org/wiki/Rete_neurale#Fondamenti_biologici

[009] http://en.wikipedia.org/wiki/Chemical_synapse

1.2 Reti Neuronal Artificiali;

[010] http://it.wikipedia.org/wiki/Rete_neurale#Fondamenti_matematico.2Finformatici

[011] http://it.wikipedia.org/wiki/Apprendimento_supervisionato

[012] http://it.wikipedia.org/wiki/Apprendimento_per_rinforzo

1.3 La Rete Feedforward;

[013] <http://spaceradioandmore.blogspot.com/2012/01/introduzione-alle-reti-neurali-ed-alla.html>

[014] http://automatica.ing.unibs.it/mco/cgsa/neurali/il_percettrone.htm

[015] <http://it.wikipedia.org/wiki/Percettrone>

[016] <http://www.cs.nott.ac.uk/~gkx/courses/g5ai/006neuralnetworks/neural-networks.htm>

[017] http://www.irccsdebellis.it/html/reti_neurali/teoria_delle_reti_neurali5.htm

[018] http://it.wikipedia.org/wiki/Rete_neurale#Funzionamento_di_una_rete_neurale_feedforward

1.4 L' algoritmo di Backpropagation;

[019] http://it.wikipedia.org/wiki/Rete_neurale#Storia

[020] http://it.wikipedia.org/wiki/Rete_neurale#Algoritmo_di_backpropagation

1.5 Intelligenza Artificiale e Informatica Musicale : Simbolismo e Connesionismo a confronto;

[021] <http://www.treccani.it/enciclopedia/connessionismo/>

[022] http://it.wikipedia.org/wiki/Connessionismo#Simbolismo_e_Connessionismo

[023] “Symbolic AI Versus Connectionism in Music Research”, Petri Toiviainen, Department of Musicology University of Jyväskylä, 2000

2.0 Percezione Sonora e Musica;

2.1 Percezione Uditiva Umana;

[024] <http://elite.polito.it/files/courses/01OQM/slide/percezione-sensoriale.pdf>

[025] <http://it.wikipedia.org/wiki/Psicoacustica>

2.2 Note e Timbro;

[026] “Neural Machines for Music Recognition”, Leo Pape, Utrecht University, March 26, 2006

[027] http://fisicaondemusica.unimore.it/Percezione_del_timbro.html

2.3 Armonia, Ritmo e Melodia;

[028] <http://it.wikipedia.org/wiki/Armonia>

[029] <http://it.wikipedia.org/wiki/Ritmo>

[030] <http://it.wikipedia.org/wiki/Melodia>

[031] [http://it.wikipedia.org/wiki/Intervallo_\(musica\)](http://it.wikipedia.org/wiki/Intervallo_(musica))

[032] http://it.wikipedia.org/wiki/Modo_musicale

2.4 Illusioni Sonore;

[033] http://fisicaondemusica.unimore.it/Effetti_e_illusioni_acustiche.html

3.0 Classificazione di File Audio Mediante Reti Neuronali;

3.1 Classificazione Audio ed Intelligenza Artificiale;

[034] “Classificazione File Audio”, Emanuele Lattanzi, Sistemi Informativi Multimediali, Università di Urbino

[035] “Music Genre Classification Systems- A Computational Approach”, Peter Ahrendt, 2006

[036] “Music Classification using Neural Networks”, Paul Scott, Stanford University, 2001

3.3 Criteri di Classificazione;

[037] “jSymbolic: A Feature Extractor for MIDI Files”, Cory McKay and Ichiro Fujinaga Music Technology Area, Schulich School of Music, McGill University

[038] <http://it.wikipedia.org/wiki/Cepstrum>

[039] “Classifying Music Based on Frequency Content and Audio Data”, Craig Dennis

[040] <http://marsyas.info/>

4.0 MIDI;

4.1 Introduzione al MIDI e alla Rappresentazione Simbolica dell'Informazione Musicale;

[041] http://it.wikipedia.org/wiki/Musical_Instrument_Digital_Interface
e

[042] “Fare musica con il PC”, Pier Calderan, Luca Barbieri, 2004