

RETI NEURALI PER IL CONTROLLO DELLE DEVIAZIONI TEMPORALI NELL'ESECUZIONE MUSICALE.

R. Bresin, G. De Poli, A. Vidolin

C.S.C. - D.E.I.
Università di Padova
via Gradenigo 6/a, I - 35131 Padova
E.mail ADTPOLI@ipduniv.bitnet
tel. +39 (0)49 8287631 - Fax +39 (0) 49 8287699

ABSTRACT

Musicians' skill has an important role in music performance. What to do with computer music performance? In this paper a connexionist approach to this problem is proposed.

An hybrid system for computer music automatic performance is showed, in which two sets of symbolic and sub-symbolic rules act in cooperation.

Some models of neural nets to calculate time and loudness deviations are discussed.

An application of the system to performe the output of an automatic source of melodies is described.

INTRODUZIONE

I musicisti, a seconda del tipo di strumento, effettuano variazioni nelle intensità, nelle durate e nei timbri delle note, poichè la notazione tradizionale non è sufficientemente precisa riguardo alle intenzioni del compositore. Infatti, se la musica scritta sul pentagramma fosse interpretata seguendo alla lettera le indicazioni metronometriche, risulterebbe estremamente meccanica ed asettica all'orecchio dell'ascoltatore. Sono proprio le deviazioni delle durate delle note rispetto alla loro durata nominale indicata dal metronomo, a rendere un'esecuzione di un interprete musicale diversa da quella di un altro. Oltre alle durate una grande importanza nell'interpretazione è costituita anche dalle variazioni di sonorità, e quindi di pressione sonora: immaginiamo infatti se un pianista, ad esempio, eseguisse una sonata di Beethoven suonando tutte le note con la stessa intensità, l'esecuzione risulterebbe estremamente piatta. I compositori danno delle indicazioni, con una simbologia opportuna, sulle variazioni di sonorità da effettuare da parte dell'interprete, le quali a loro volta non sono seguite alla lettera, così come le indicazioni metronometriche, ma sono "interpretate" secondo il gusto, la cultura e l'abilità dell'esecutore. In sostanza sono le variazioni metronometriche e quelle di sonorità a caratterizzare maggiormente lo stile interpretativo e l'abilità di un esecutore rispetto ad un altro.

In particolare Repp è stato il primo ad analizzare un campione statisticamente rappresentativo di registrazioni commerciali dello stesso brano (Terzo movimento della Sonata per pianoforte op.31 No.3 di Beethoven) eseguito da 19 pianisti di fama mondiale. Da tale analisi si nota che in tutti esecutori le durate dei quarti di ogni battuta hanno subito delle deviazioni, sia in incremento che in decremento, rispetto alle durate metronometriche nominali.

Il problema dell'interpretazione è cruciale nella computer music, dove l'esecuzione è lasciata al computer. Infatti l'esecuzione ottenuta facendo una traduzione letterale delle classiche indicazioni interpretative presenti negli spartiti musicali, porta ad un'esecuzione estremamente meccanica. Si presenta quindi la necessità di formalizzare l'interpretazione così da avere un'esecuzione più "calda", più "umana".

Varie sono state le proposte. Negli anni '30 Carl Seashore e i suoi collaboratori alla Iowa

University (Gabrielsson, 1985) introdussero delle tecniche particolari per lo studio dell'esecuzione pianistica. Venne realizzato un sistema costituito da strisce di balsa incollate ai martelletti. Ognuna delle strisce aveva un foro, che, quando il martelletto veniva colpito, era attraversato da un fascio di luce, che a sua volta impressionava la pellicola di una cinepresa. Con questa tecnica, se pur empirica poichè consentiva un'accuratezza di 0.01 secondi, Seashore fece delle importanti osservazioni: fu notata, ad esempio, una differenza durata tra i quarti all'interno di una stessa battuta.

Altri ricercatori hanno cercato di estrarre delle regole interpretative dall'ascolto di esecuzioni di musicisti esperti, come Repp (1990) e Clynes. Quest'ultimo è arrivato ad ipotizzare un andamento metronometrico caratteristico per i maggiori compositori della musica classica occidentale, proponendo delle microstrutture espressive che, introdotte nelle esecuzioni mediante computer, danno all'ascoltatore l'idea sia della personalità del particolare compositore che del suo "stile di movimento" (Repp 1991).

Un'altra proposta è stata quella di De Poli, Irone e Vidolin (1990) che hanno messo a punto un sistema esperto per l'interpretazione di spartiti musicali mediante una base di dati multilivello. In questo approccio sono state utilizzate due basi di conoscenza: una musicale, che codifica l'esperienza e la fantasia dell'interprete, e una sonologica, costituita da un insieme di regole per ogni strumento, che fornisce i parametri fisici dell'esecuzione partendo dallo spartito simbolico.

Un diverso approccio è quello di costruire delle regole simboliche con il metodo di analisi mediante sintesi, proposto da Sundberg e collaboratori (Sundberg et al., 1991; Friberg, 1991). Con tale metodo si sono interrogati musicisti ed insegnanti di musica esperti, che hanno permesso così di formulare delle regole interpretative. Tali regole vengono poi applicate in esecuzione una dopo l'altra secondo l'indicazione di un musicista esperto, il quale giudica se l'effetto aggiuntivo dovuto ad un'ulteriore regola sia gradevole o meno. In pratica l'esecuzione è ottenuta dalla somma delle deviazioni temporali, sonore e timbriche ottenute dall'applicazione contemporanea di un certo numero di regole simboliche.

In questo lavoro si è pensato di utilizzare un sistema ibrido in cui vi siano regole sub-simboliche in eventuale combinazione con regole simboliche. L'idea è di combinare piccole reti controllate da regole simboliche.

REGOLE PER L'ADDESTRAMENTO DELLE RETI NEURALI PER L'ESECUZIONE DI PARTITURE.

Per l'addestramento delle reti neurali è stato utilizzato l'algoritmo di back-propagation, che prevede dei dati in input alla rete e dei dati da imporre al suo output. A questo scopo sono state utilizzate alcune delle regole proposte da Sundberg, considerando i parametri che compaiono in esse come input per la rete, e il risultato, somma di tutte le regole considerate, come output della rete stessa (cioè la deviazione temporale o di loudness, a seconda della rete).

Consideriamo prima la parte sub-simbolica del sistema, in cui si descrive come le reti vengono costruite ed addestrate.

Sundberg e i suoi collaboratori del KTH di Stoccolma (Sundberg et al., 1991) hanno proposto varie regole interpretative ottenute mediante il metodo di analisi mediante sintesi di cui si è parlato in precedenza. Lo scopo delle regole è di convertire la partitura musicale, completa di simboli di frase e di accordi, in un'esecuzione musicalmente accettabile. Le deviazioni risultanti dall'applicazione delle regole sono di tipo additivo, in modo che ogni nota può essere modificata da più regole, e le deviazioni dovuta ad ogni regola sono aggiunte successivamente ai parametri di tale nota. L'ordine con cui tali regole sono applicate non è rilevante, tranne che per le regole di sincronizzazione e quelle che interessano l'ampiezza dell'involuppo delle note, che devono essere applicate per ultime. Le regole possono essere suddivise in cinque gruppi: regole che interessano un parametro, regole che interessano più parametri, regole di intonazione, regole che agiscono sull'ampiezza dell'involuppo e regole di sincronismo. Alcune di queste regole sono di segmentazione, per enfatizzare le singole note, altre sono di raggruppamento, che aiutano l'ascoltatore a raggruppare elementi strutturali (temi, frasi, etc.), e tra quest'ultime alcune interessano la partitura ad un livello macroscopico, altre ad un livello microscopico (locale). Infine vi sono delle regole di insieme, quali quelle di sincronismo. La maggior parte delle regole include il parametro k . Esso è usato come peso della regola, per alterarne l'influenza, e dovrebbe essere

usato con lo stesso valore per tutte le regole. Il valore di default è $k = 1$, ed è appropriato quando tutte le regole sono applicate. Se una regola è applicata isolatamente, allora, secondo Friberg (1991), è meglio utilizzare valori leggermente più elevati di k per produrre delle variazioni udibili. Inoltre, per ottenere il miglior risultato, per un dato brano musicale, Friberg consiglia di utilizzare un valore di k diverso per ogni regola.

E' possibile quindi utilizzare diversi valori di k per ottenere differenti interpretazioni dello stesso brano musicale, come vedremo più avanti.

Fra tutte le regole proposte da Sundberg e collaboratori abbiamo scelto quelle che interessassero contemporaneamente il minor numero di parametri, per poter lavorare con reti neurali di piccole dimensioni, e che allo stesso tempo fossero tra le più significative dal punto di vista dell'esecuzione musicale. Alla fine si sono considerate 6 regole, che agiscono sulle deviazioni temporali o su quelle di loudness o su entrambi:

- 1 - contrasto di durata (durational contrast)
- 2 - doppia durata (double duration)
- 3 - carico melodico (melodic charge)
- 4 - intensità elevata (high loud)
- 5 - durata della nota in un salto (leap tone duration)
- 6 - salita più veloce (faster uphill)

Le prime quattro regole scelte sono di segmentazione in quanto agiscono isolatamente sulla singola nota, mentre le ultime due sono di raggruppamento.

In particolare le regole 1, 2, 3, 5 e 6 interessano le variazioni di durata della singola nota, e fra esse le regole 1 e 3 insieme alla 4 agiscono anche sulla variazione del loudness, sempre della singola nota.

CODIFICA DEI DATI IN INGRESSO ED USCITA NELLE RETI NEURALI E SCELTA DEI PARAMETRI PER L'APPRENDIMENTO.

Un aspetto fondamentale nell'uso delle reti neurali è il modo in cui si codificano i dati in ingresso e in uscita alla rete. A seconda dei modelli di funzione non lineare utilizzata dalla rete, l'input e l'output devono presentare valori compresi tra 0 e 1 oppure tra -1 e +1. Ad esempio quando la funzione non lineare è la sigmoide, come nel nostro caso, l'input e l'output devono avere valori compresi nell'intervallo [0, 1], mentre se la funzione non lineare è la tangente iperbolica l'intervallo di ammissibilità per i valori è [-1, +1]. In entrambi i casi si considerano compresi gli estremi. Tuttavia, per quanto riguarda l'output, è meglio limitare l'intervallo dei valori ammissibili per non rischiare di mandare in saturazione la rete facendola lavorare in corrispondenza del "ginocchio" della funzione non lineare: ad esempio Rumelhart e McClelland (1988b) consigliano per l'output, nel caso della sigmoide, l'intervallo di valori [0.1, 0.9].

Una volta stabilito il range di valori ammissibili in ingresso ed in uscita dalla rete, è di fondamentale importanza il modo in cui vengono scelti e codificati i dati a disposizione cercando di ricondurli in tale range, ed è proprio da questa codifica e da una sua corretta interpretazione che dipende la riuscita o meno dell'addestramento di una rete neurale.

In primo luogo bisogna assicurarsi che fra tutti i pattern di training utilizzati non ce ne siano di uguali che diano output diversi, perché porterebbero la rete a fornire in output un valore medio anziché quello desiderato. Quindi bisognerà scegliere i dati da dare in input e la loro rispettiva codifica in modo da evitare ogni ambiguità.

Nel caso specifico di questa ricerca, consideriamo ora la prima rete neurale artificiale realizzata, riportata in figura 1. In essa si vede che vi sono quattro neuroni di input e uno di output.

Il primo neurone porta la scritta "Durata Nominale", che sta ad indicare che esso riceve in input la durata della nota così come risulta dalle indicazioni metronometriche della partitura originale. Poiché il neurone accetta in ingresso solo dati tra 0 ed 1 e dal momento che le durate utilizzate, per esigenze di calcolo, sono espresse in millisecondi ed essendo la regola del contrasto di durata l'unica, tra le regole scelte, in cui compare la durata nominale della nota in esame e non avendo effetto per note con durata superiore ai 600 ms, si è pensato di dividere per 1000 le durate nominali. Nel caso di note di durata superiore ai 1000 ms, esse vengono forzate ad avere durata 1000 (solamente in ingresso alla rete).

Il secondo neurone di input è quello che tiene conto del carico melodico. In ingresso a tale

neurone viene dato il valore assoluto del carico melodico della nota in esame diviso per il valore assoluto del massimo carico melodico possibile, che è pari a 6.5: $\frac{|C_{mel}|}{6.5}$, per cui l'input sarà sempre tra 0 (nel caso di carico melodico $C_{mel} = 0$) e 1 (nel caso di $C_{mel} = -6.5$).

Il terzo ed il quarto neurone di input servono entrambi a codificare le informazioni necessarie all'apprendimento della regola di salto. Si sono scelti due neuroni per tale regola dopo varie prove con un solo neurone, in cui ci si è accorti che si incappava in casi di forte ambiguità. Ad esempio salti ascendenti con note di stessa durata e stesso carico melodico davano output diversi. Si è vista allora la necessità di introdurre un ulteriore neurone di input, che tenesse conto del numero di semitoni presenti tra le due note formanti il salto. In particolare, per quanto riguarda il terzo neurone, sono state adottate le seguenti codifiche:

- a) input = 0.1: nel caso che la nota in esame sia la prima del salto;
- b) input = 0.5: nel caso che la nota in esame non soddisfi alle specifiche della regola del salto;
- c) input = 0.7: nel caso che la nota in esame sia la nota di arrivo in un salto discendente;
- d) input = 0.9: nel caso che la nota in esame sia la nota di arrivo in un salto ascendente.

Tali valori sono stati ottenuti considerando i coefficienti moltiplicativi presenti nelle regole di salto. Si è osservato che il coefficiente per la prima nota è sempre -4.2 e che quello per la nota di arrivo nel caso di salto ascendente è +4.2, che è il più grande tra i due coefficienti positivi. Allora è stata realizzata una funzione lineare che mappa l'intervallo [-4.2, 4.2] nell'intervallo [0.1, 0.9], ottenendo così che lo 0 dell'intervallo di partenza (che indica l'assenza di salto, poichè come coefficiente moltiplicativo darebbe una deviazione nulla) viene mappato nel valore 0.5 dell'intervallo [0.1, 0.9].

Il valore in ingresso al quarto neurone di input di figura 1 indica il numero di semitoni tra le due note del salto. Tale valore, dovendo essere compreso tra 0 e 1, viene ottenuto dividendo il numero di semitoni per 9 e se il quoziente risulta maggiore di 1, esso viene forzato a valere 1. Cioè si è pensato che nei salti con numero di semitoni superiore a 9, la rete si deve comportare come nel caso di nove semitoni. Infatti, se non si agisse in questo modo, si potrebbero ottenere delle deviazioni temporali esagerate, poichè la radice quadrata del numero di semitoni interviene direttamente come fattore moltiplicativo nella regola del salto (considerando anche che nella moltiplicazione interviene il parametro K, che, come si vedrà in seguito, è stato scelto con valori da 1 a 9).

Anche per la codifica dei valori da insegnare alla rete durante la fase di apprendimento si è utilizzata una funzione lineare del tipo di quella del caso del salto. L'inversione di tale codifica permetterà poi di decodificare l'output della rete dopo la fase di addestramento.

In particolare l'output che viene inizialmente insegnato alla rete, mediante l'algoritmo di back-propagation, è ottenuto applicando le regole simboliche scelte allo spartito in esame (nel nostro caso il Minuetto di Beethoven). Le regole forniscono per ogni nota la deviazione temporale corrispondente ed è proprio questa che viene insegnata alla rete. Pensando di avere al massimo deviazioni di 100 ms, queste vengono divise appunto per 100, ottenendo così valori compresi tra -1 e 1, potendo esserci sia deviazioni positive che negative. Questi valori vengono poi mappati nell'intervallo [0.1, 0.9] dove allo 0 (corrispondente a deviazione nulla) dell'intervallo [-1, 1] corrisponde il valore 0.5 nell'intervallo [0.1, 0.9]: cioè i valori tra -1 e 0 sono stati mappati in [0.1, 0.5] e quelli tra 0 e +1 in [0.5, 0.9].

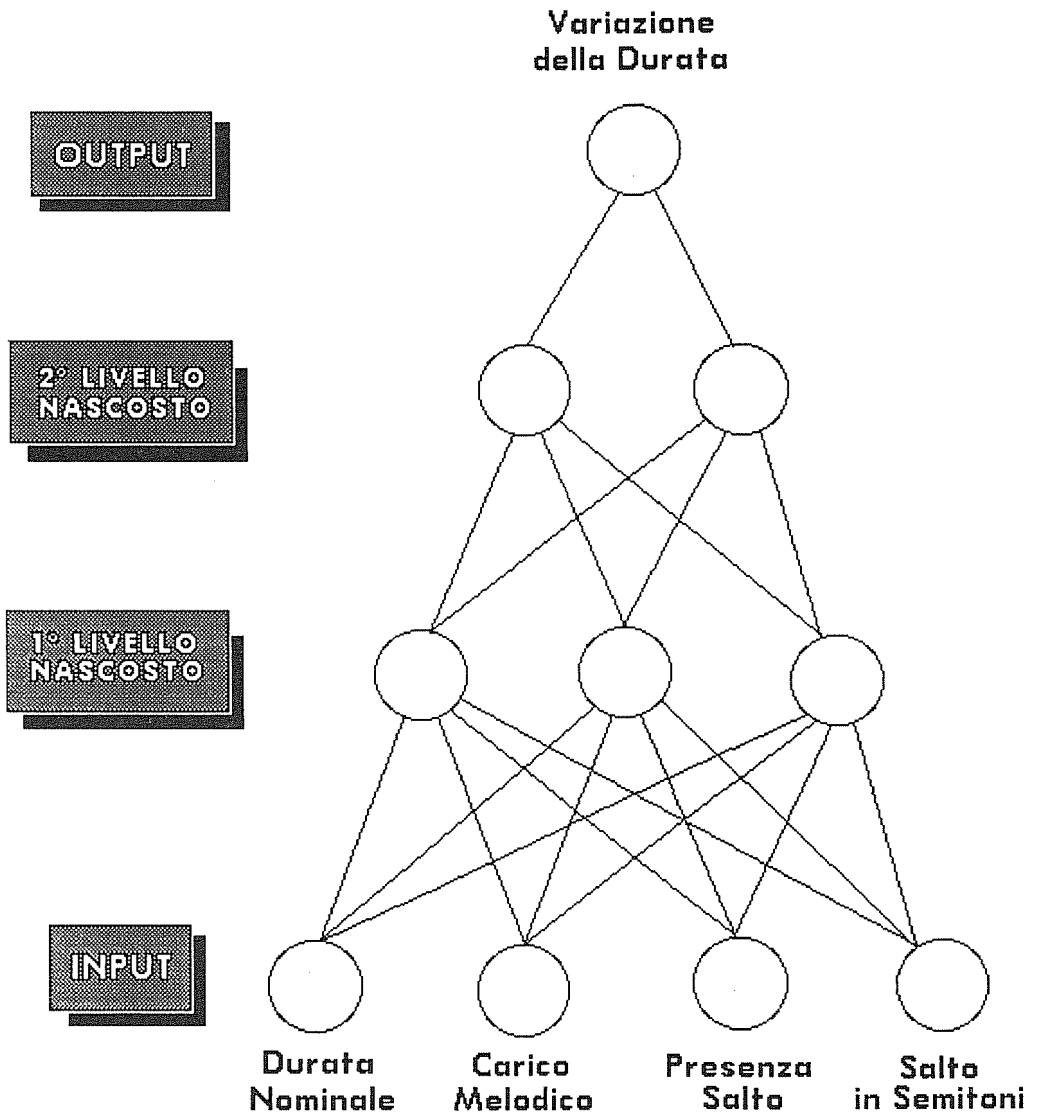


figura 1: Rete neurale per il calcolo delle deviazioni temporali

APPRENDIMENTO DELLE REGOLE

Per insegnare le regole di Sundberg e collaboratori scelte per questo lavoro, è stato considerato il III movimento della sonata per pianoforte op. 31 No. 3 di Beethoven. Si è scelto questo brano in quanto già oggetto di studio da parte di Repp (1990), potendo così confrontare ed insegnare i risultati del presente lavoro con quelli del ricercatore americano.

Il III movimento della sonata in oggetto è così strutturato: Minuetto - Trio - Minuetto. Si veda l'appendice B.

Lo studio in questa tesi è stato focalizzato sulla prima delle due parti di Minuetto. Come esempio per l'addestramento ho preso 16 note, prese sia dal primo ritornello che dal secondo, in modo da fornire il maggior numero di esempi significativi alla rete. In particolare le note scelte sono 16, e corrispondono ai seguenti casi:

- 1 salto ascendente
- 1 salto discendente (di ampiezza diversa da quella del salto ascendente)
- 12 note di durate anche uguali, ma con carichi melodici diversi

Verranno ora discussi alcuni grafici ottenuti testando le reti neurali già addestrate con dei particolari pattern di input (Rumelhart e McClelland 1988b), che mettono in evidenza, come si vedrà in seguito, come la rete abbia effettivamente appreso le regole insegnate (nel caso particolare qui riportato si tratta della regola del salto).

La struttura delle prime reti addestrate è quella riportata in figura 1, in cui vi sono 4 neuroni di input, 1 di output e 2 livelli nascosti. La scelta di tale struttura è motivata dal fatto che con due livelli nascosti, anziché uno, la rete apprende più velocemente, avendo più gradi di libertà a disposizione per adattare la configurazione dei pesi sulle sinapsi (rami) a seconda del particolare pattern di input. Chiaramente un apprendimento veloce implica anche una opportuna scelta del *momentum* e del *learning rate*, che dopo alcune prove sono stati settati rispettivamente a 0.99 e 0.25.

Per l'addestramento di una rete con tale struttura, utilizzando un insieme di 16 pattern di training come quello descritto in precedenza, sono stati necessari circa 15 minuti di tempo di calcolo di un personal computer con microprocessore 80386 a 33 MHz. Questo è un dato confortante, pensando, come si vedrà e ascolterà dai risultati ottenuti, che uno studente principiante arriva ai risultati interpretativi ottenuti con queste semplici reti almeno dopo un anno di studi.

Risposte della rete al variare del numero dei semitoni tra le note in un salto.

Interessante è vedere come varia l'output quando, con durata = 0 e melodic charge = 0, si impone salto = 0.1, oppure salto = 0.7 o infine salto = 0.9, facendo variare di volta in volta il numero dei semitoni, che viene fatto variare il numero di semitoni tra le due note del salto da 0 a 9 (la scelta di 9 come numero massimo dei semitoni è dovuta al fatto che se si presenta un salto con numero di semitoni maggiore a 9, questo viene forzato a 9 in input alla rete in modo da non avere delle deviazioni temporali esagerate, che invece si avrebbero per valori elevati del numero dei semitoni). Questo è stato fatto per $K = 9$ e i risultati sono riportati nelle figure 2, 3 e 4. Per salto = 0.1 e salto = 0.9 si vede come l'output abbia andamento rispettivamente decrescente e crescente all'aumentare del numero dei semitoni. Per salto = 0.7 la risposta della rete segue l'andamento crescente previsto dalla regola di Sundberg fino a circa 7 semitoni. Poi la risposta comincia a decrescere, mantenendo comunque una deviazione positiva. Questi andamenti corrispondono in maniera soddisfacente alle relative regole (dove ΔDR indica la variazione di durata e il pedice n indica la prima nota del salto):

nota di partenza: $\Delta DR = -4.2 * \sqrt[2]{\Delta N * K}$ [ms]

nota di arrivo per un salto ascendente: $\Delta DR = 4.2 * \sqrt[2]{\Delta N * K}$ [ms]

nota di arrivo per un salto discendente: $\Delta DR = 2.4 * \sqrt[2]{\Delta N * K}$ [ms]

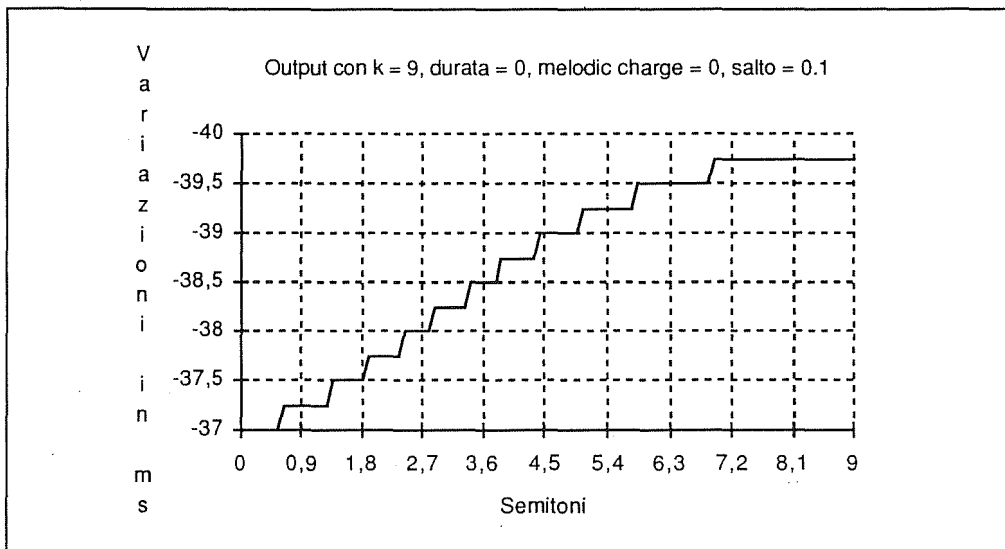


figura 2: Nota di partenza del salto

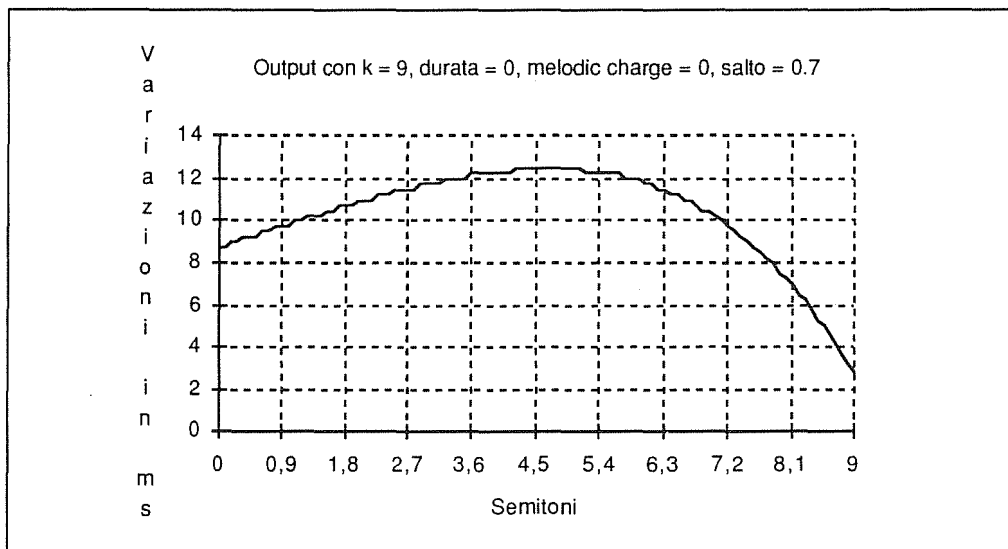


figura 3: Nota di arrivo nel salto discendente

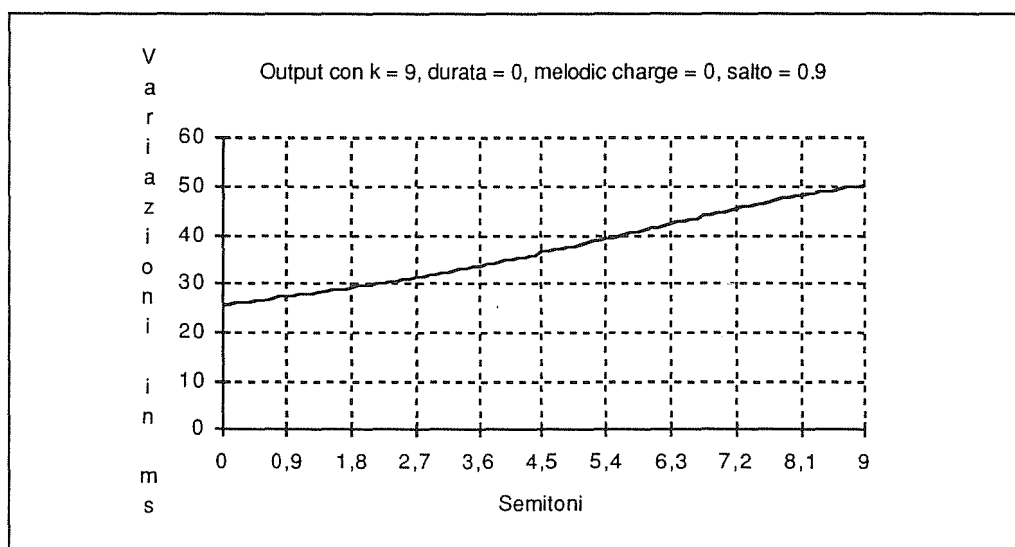


figura 4: Nota di arrivo nel salto ascendente

CONFRONTO TRA LE DEVIAZIONI TEMPORALI OTTENUTE COME OUTPUT DELLA RETE E COME OUTPUT DESIDERATO, NEL CASO DELL'ESECUZIONE DEL MINUETTO.

Dopo l'analisi delle uscite delle tre reti addestrate per $K = 1$, $K = 5$ e $K = 9$ è interessante sottolineare che nell'esecuzione dell'intero minuetto di Beethoven le reti hanno risposto benissimo per note di durate che non erano state apprese durante la fase di learning. Inoltre, soprattutto per la rete addestrata con $K = 9$, la deviazione temporale ha mostrato un andamento abbastanza diverso da quello previsto dalle regole di Sundberg in corrispondenza dei salti. In particolare lì dove, in corrispondenza della nota di partenza del salto, le regole di Sundberg forniscono una deviazione temporale negativa piuttosto accentuata, la rete tende a diminuire, in valore assoluto, tale decremento. Dove invece le regole, sempre in corrispondenza della nota di partenza del salto, darebbero un incremento della durata della nota, in seguito alla concomitanza dei contributi con deviazioni positive dovuti alle regole diverse da quella del salto, la rete dà un decremento o comunque un incremento minore di quello ottenibile con le regole simboliche. Inoltre, nel caso di prima nota del salto, ove il decremento sia di piccola entità, questo viene leggermente aumentato in valore assoluto.

Invece, in corrispondenza della nota di arrivo di un salto ascendente, la deviazione fornita dalla rete è uguale o leggermente superiore a quella che si otterrebbe con l'applicazione diretta delle regole di Sundberg. Nel caso del salto discendente, invece, la deviazione temporale fornita dalla rete è sempre leggermente superiore a quella che si otterrebbe dalle regole simboliche di Sundberg.

Quindi, in sostanza, il risultato più interessante è quello che interessa la nota di partenza in un salto tra due note. Infatti dai risultati ottenuti è chiaro che la non linearità intrinseca nella rete neurale operi con un'azione di "smussamento" fornendo dei risultati migliori di quelli ottenibili mediante un'applicazione additiva, spesso esagerata, delle regole simboliche di Sundberg. Inoltre, quando le regole simboliche forniscono un incremento della durata della prima nota del salto, la rete neurale, cioè le regole sub-simboliche, tendono a fornire un decremento o un incremento pressoché nullo, indicando quindi che il "concetto di decremento della durata in corrispondenza alla prima nota di un salto" è stato appreso molto bene dalla rete.

SISTEMA IBRIDO

In analogia con lo schema di figura 6 (Hubel, 1975) si può pensare di rappresentare il sistema utilizzato per il presente lavoro come in figura 5, dove tutto ciò che sta a monte dei livelli nascosti della rete neurale corrisponde ai livelli da A a C di figura 6 e tutto ciò che sta a valle è analogo ai livelli da X a Y della stessa figura.

In pratica potremmo pensare che il livello dei neuroni di input corrisponde ai nervi ottici del pianista stimolati dalla lettura della partitura (causa), mentre il livello di output corrisponde all'effetto, quindi all'azione meccanica, corrispondente al particolare input.

La partitura può essere acquisita mediante un sintetizzatore collegato ad un personal computer tramite interfaccia MIDI, oppure con un editor musicale o mediante un opportuno file dati. A questo punto la partitura viene sia analizzata dalle due regole simboliche utilizzate in questo lavoro che produrranno un particolare output (OUT_2 in figura 5), che codificata in maniera opportuna, così da poter essere "letta" dalla rete neurale. Quest'ultima, in seguito all'applicazione delle regole sub-simboliche, fornirà a sua volta un output che, opportunamente decodificato (OUT_1 in figura 5), verrà sommato a quello fornito dalle regole simboliche. Il risultato della sommatoria verrà poi applicato alla partitura originale per dare la partitura "esecutiva", cioè la partitura ottenuta con l'applicazione delle regole. Quindi la nuova partitura viene inviata, attraverso l'interfaccia MIDI, al sintetizzatore per il suo ascolto. Inoltre, un'analisi mediante regole simboliche della partitura permette di determinare (nel caso del Minuetto in esame) se si sta eseguendo o no il ritornello, facendo agire, a seconda dei casi (freccia tratteggiata), una rete oppure l'altra (addestrate ad esempio con un diverso k, o con il diverso stile che il pianista adotta a seconda dei casi) oppure scegliendo il valore di k (come si è visto nelle reti con 5 neuroni di input).

Analizziamo il sistema nelle sue parti principali.

Regole simboliche.

Le regole simboliche utilizzate in questo sistema sono due ed agiscono entrambe sulla variazione della durata della singola nota. Esse sono le regole 2 e 6 riportate nel capitolo 3. Si è pensato di non insegnare queste regole alla rete neurale perchè estremamente semplici e di applicazione immediata.

Regole sub-simboliche.

Per l'applicazione di queste regole si sono utilizzate in un primo momento due reti neurali: una per ottenere le deviazioni temporali e l'altra per le deviazioni di loudness (trasformate in una variazione di key velocity).

La rete utilizzata per insegnare le regole di deviazione temporale è quella riportata in figura 7. L'addestramento utilizzato è stato quello mediante l'algoritmo di back propagation, in cui al neurone di output è stato imposto, di volta in volta, il valore ottenuto mediante l'applicazione delle tre regole adoperate.

Al primo neurone di input viene assegnato il valore della durata effettiva della nota così come riportato nella partitura originale.

Al secondo neurone di input viene assegnato il valore del carico melodico della nota in esame, così che essa viene inquadrata nel particolare contesto armonico.

Gli ultimi due neuroni di ingresso sono stati utilizzati per insegnare alla rete la regola del salto. In particolare il primo dei due viene utilizzato per indicare se la nota in esame è quella di partenza o di arrivo nel salto, e nel secondo caso si distingue con un'opportuna codifica se si tratta di un salto ascendente o discendente. Il secondo di questi due neuroni indica la lunghezza del salto in semitoni.

Tutti i valori in ingresso sono opportunamente scalati tra 0 e 1, poichè la funzione non lineare adottata dalla rete utilizzata è la sigmoide.

Il nodo di output fornisce la deviazione della durata per la particolare nota in esame. Anche il valore dell'output è scalato, questa volta tra 0.1 e 0.9 per evitare di mandare in saturazione la rete: al valore 0.5 corrisponde una deviazione nulla, per valori compresi tra 0.1 e 0.5 si ha un accorciamento della durata nominale e per valori tra 0.5 e 0.9 un aumento di essa.

Sono state addestrate tre reti neurali come questa, utilizzando per ognuna un diverso valore del

parametro k di pesatura delle regole: in particolare si sono usati i valori $k = 1$, $k = 5$ e $k = 9$. Per $k = 1$ si ottengono delle variazioni difficili da udire, mentre con $k = 9$, essendo maggiormente amplificate le regole, si hanno delle deviazioni temporali molto marcate.

In un secondo momento è stata realizzata un'altra rete neurale con 5 neuroni di input anziché 4 come in precedenza. I primi 4 nodi ricevono gli stessi input della rete precedente, mentre al 5° viene assegnato il valore del parametro k . In questo modo, quando si sta effettuando un'esecuzione in tempo reale, è possibile, variando il valore di k , ottenere deviazioni temporali più o meno accentuate, pensando così di passare da un'interpretazione in stile barocco ($k = 1$) ad una in stile romantico ($k = 9$).

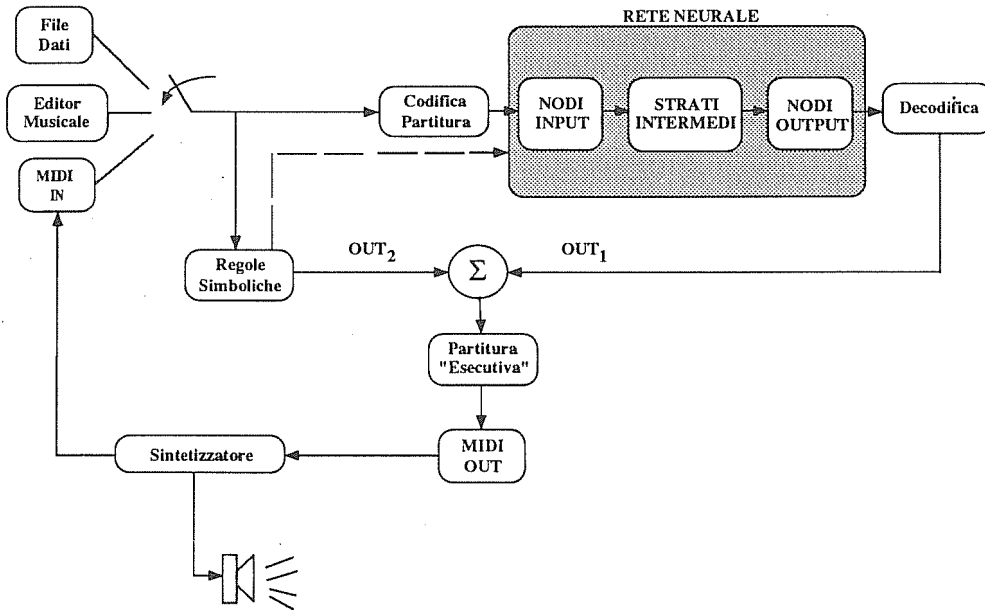


figura 5: Schema del sistema ibrido

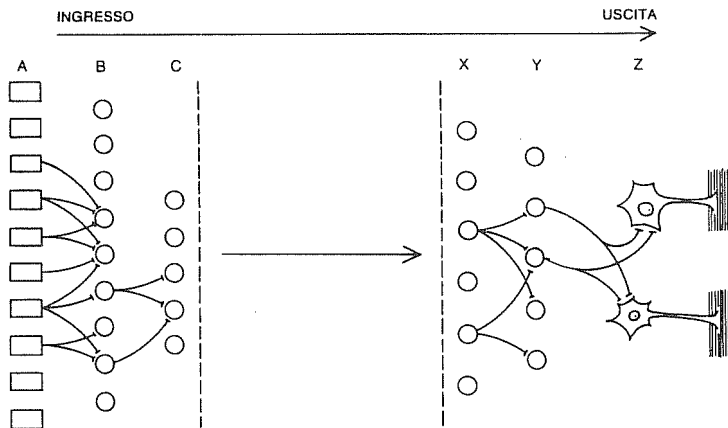


figura 6: Schema del cervello

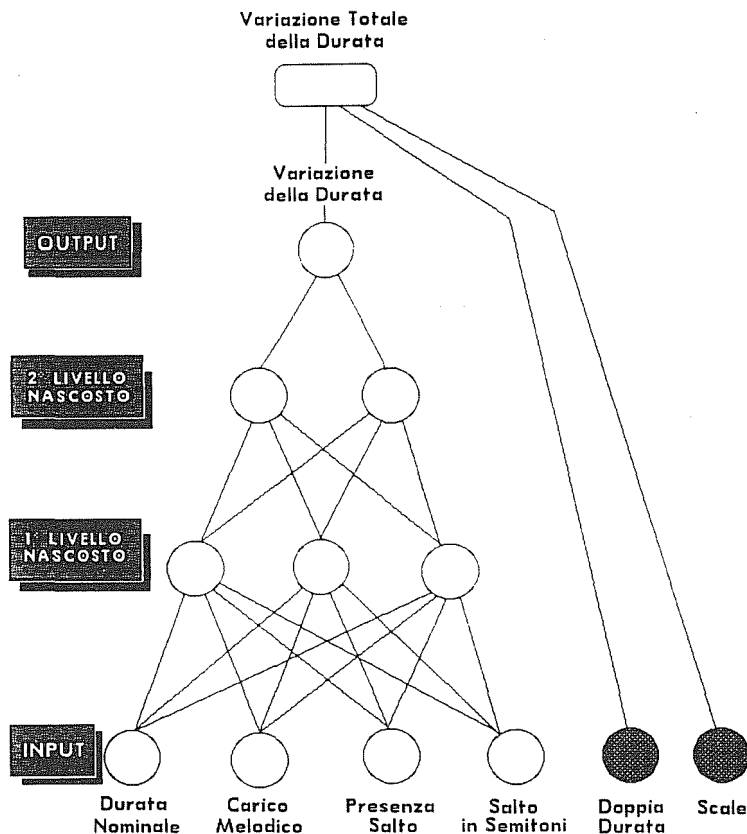


figura 7: Rete ibrida

APPENDIMENTO DI UNO STILE INTERPRETATIVO.

Un'applicazione molto interessante è quella di riuscire a realizzare una rete neurale in grado di imparare lo stile interpretativo di un pianista affermato, così da poter in seguito eseguire automaticamente delle partiture con il suo stile.

Per l'addestramento si è utilizzato il III movimento della Sonata per pianoforte op. 31 No. 3 di Beethoven. La scelta è caduta su questo brano per eseguire dei confronti con i risultati del lavoro di Repp (1990), basato sulla stessa composizione. In particolare è stata analizzata la melodia di questo pezzo, eseguita al sintetizzatore dal pianista Piero Pittari, il quale ha eseguito più volte la partitura, scegliendo alla fine quelle che egli considerava le migliori esecuzioni. Dall'analisi delle variazioni metronometriche dei singoli quarti di battuta (figure 8 e 9), si nota che un overlap abbastanza buono con i dati riportati da Repp, ottenuti dallo studio delle esecuzioni (su disco o nastro) di diciannove pianisti famosi. Si vedano a tal proposito le figure dalla 8 alla 10. L'andamento a "V" di alcune battute manifesta quello che Clynes chiama il "Beethoven pulse", in cui il II quarto della battuta è più veloce del I e il III quarto è più lento del primo, e quindi

possiamo dire che l'esecuzione a nostra disposizione è abbastanza beethoveniana.

Una volta in possesso di tutti i dati MIDI relativi all'esecuzione di Pittari, si è pensato di realizzare una rete neurale che apprendesse il suo stile interpretativo, con la possibilità, in seguito, di interpretare altre partiture con lo stile di Pittari. Dopo alcune prove condotte utilizzando la stessa rete impiegata per l'apprendimento delle regole di Sundberg e collaboratori, ci si è accorti che tale rete non era sufficiente allo scopo. L'inadeguatezza era dovuta al fatto che erano troppo pochi i parametri in ingresso ai nodi della rete, insufficienti cioè a definire l'interpretazione di un pianista professionista. Si è allora introdotto un ulteriore neurone di input per tener conto dei margini delle legature di frase, in cui si verificano solitamente rallentamenti e/o respiri, codificando se siamo in presenza di un inizio di frase o di un accelerando (valore = 0.1), in un fine di frase o di un rallentando (valore = 0.9) o in nessuno di questi casi (valore = 0.5). I valori scelti si spiegano pensando che dove non ci sia né rallentando, né accelerando non si devono ottenere microvariazioni e quindi si codifica questa situazione con il valore 0.5. Invece il caso di un accelerando, corrispondente ad una deviazione temporale negativa, viene codificato con 0.1 e quello di rallentando corrispondente ad una deviazione positiva con 0.9. Inoltre si sono considerate due reti diverse nel caso si consideri il I o il II ritornello, poichè le deviazioni temporali apportate dal pianista sono decisamente diverse nei due casi, questo per dare maggior "movimento" all'esecuzione. L'addestramento della rete è stato fatto sempre mediante l'algoritmo di back propagation, dove l'output insegnato è la deviazione temporale percentuale della singola nota ottenuta facendo il rapporto tra la durata risultante dall'esecuzione e la durata nominale riportata dalla partitura. Tra le deviazioni percentuali così ottenute sono state scartate quelle con valore troppo elevato, che si sono verificate prevalentemente all'inizio e alla fine del brano e alla fine dei ritornelli.

Il risultato ottenuto è buono: proponendo l'intero minuetto eseguito dalla rete addestrata ad alcuni ascoltatori, questi hanno riconosciuto alcune caratteristiche stilistiche del pianista Pittari.

In figura 10 è rappresentato l'andamento dei quarti delle battute del minuetto eseguito dalla rete addestrata con lo stile del pianista. Si può osservare, che ci sono ancora i caratteristici andamenti a "V" in almeno 20 battute, non sempre corrispondenti a quelli dell'esecuzione originale, ma che comunque danno un andamento abbastanza beethoveniano (secondo Clynes) dell'esecuzione mediante rete neurale addestrata. In ogni caso, a livello microscopico, è importante notare come sia preservato lo stile del pianista (vedi figura 9), dato da deviazioni temporali marcate tra il terzo quarto della battuta e il primo quarto della battuta successiva.

Se confrontiamo questo grafico con quello dell'andamento dei quarti del minuetto eseguito dal sistema ibrido (con $K = 9$: valore che dà le maggiori deviazioni), si nota che il primo grafico dà decisamente un'esecuzione con deviazioni temporali maggiori. Questo è dovuto soprattutto al fatto che per il sistema ibrido sono state considerate poche regole e quasi tutte di segmentazione e che, in ogni caso, non possono tener conto di quelle implicazioni culturali ed emotive intrinseche nell'esecuzione di un pianista. Per cui i risultati ottenuti sono da considerarsi interessanti.

Per ottenere risultati ancora migliori, per l'apprendimento dello stile di un pianista, si potrebbe utilizzare un modello di rete più sofisticato, come quello della rete neurale ecologica (Parisi, 1990): cioè una rete che tenga conto del fatto che gli stimoli provenienti dall'ambiente (input della rete) nel ciclo N sono funzione delle caratteristiche dell'ambiente ma anche dell'azione della rete stessa (output della rete) nel ciclo $N-1$. Si potrebbe allora pensare di utilizzare una rete con gli stessi input di quella dell'esperimento precedente con in più un ulteriore nodo di input, che tiene conto dell'output della rete al passo precedente. Si ottiene così una rete neurale con feed-back.

In quest'ultimo caso si può pensare che i neuroni di input rappresentino non solo i nervi ottici, cioè gli occhi del pianista durante la lettura della partitura, ma anche, con il nuovo neurone di input, i nervi acustici, cioè il pianista si ascolta durante l'esecuzione e quindi agisce di conseguenza.

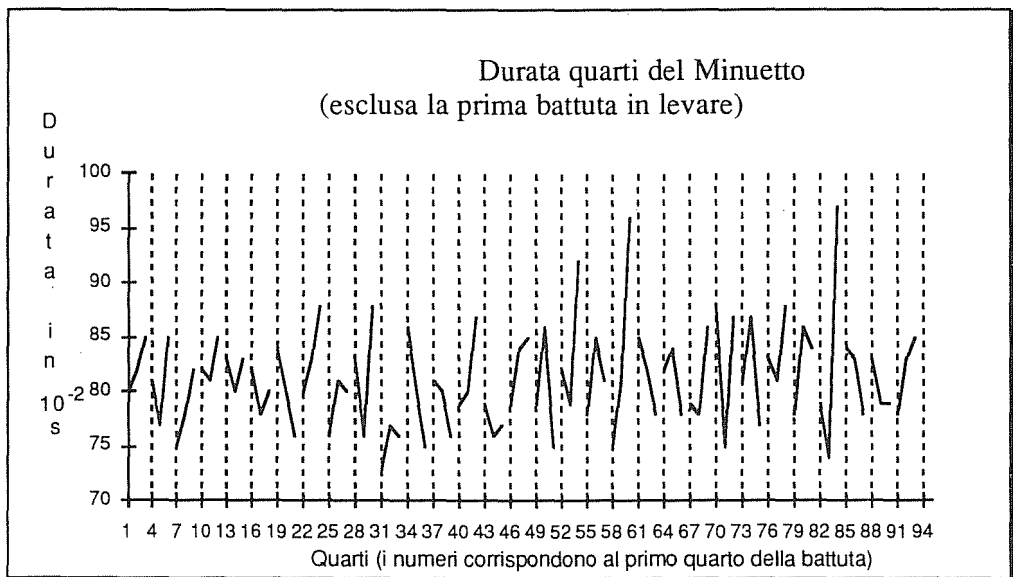


figura 8: Minuetto eseguito dal pianista

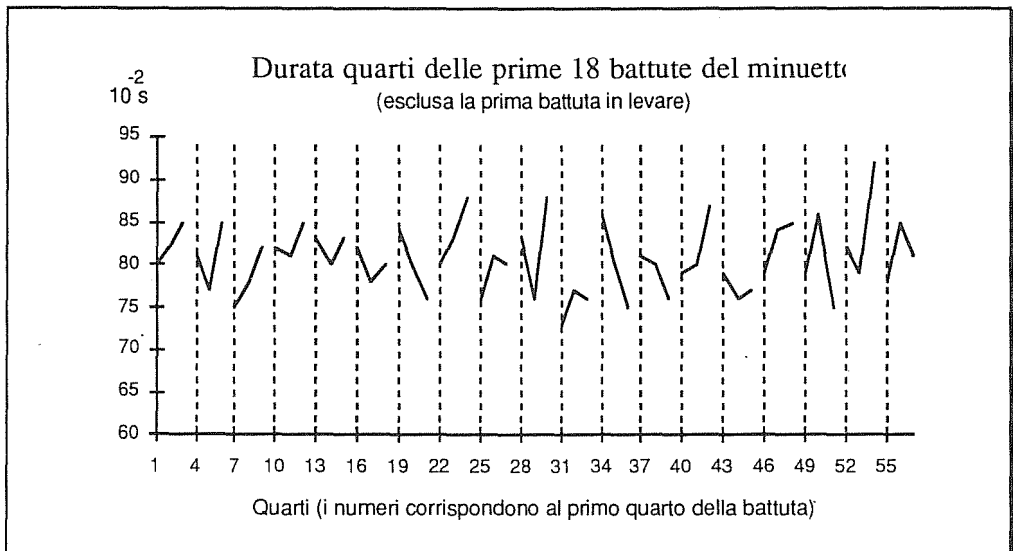


figura 9: Minuetto eseguito dal pianista

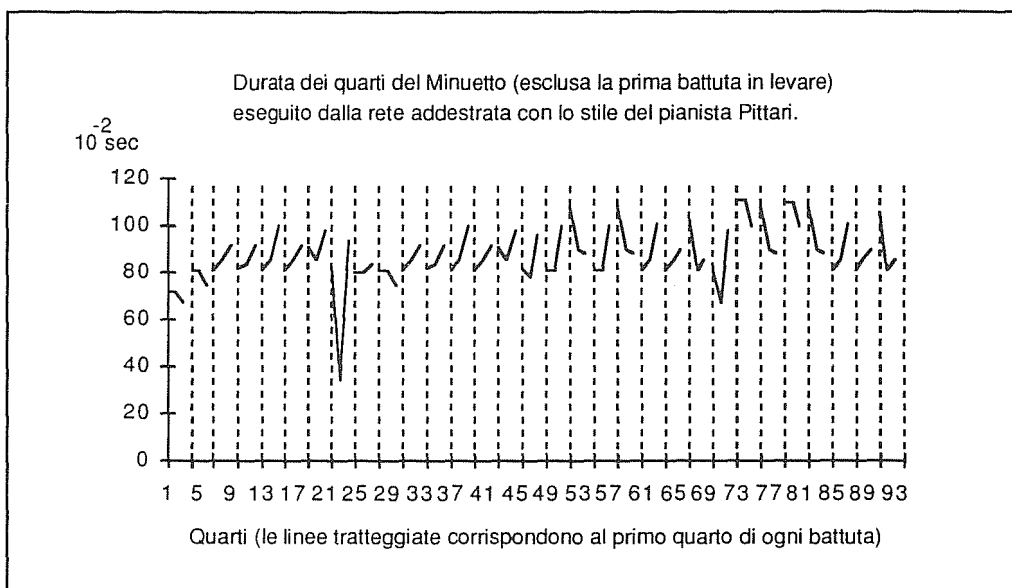


figura 10

CONCLUSIONI

I buoni risultati ottenuti in questo lavoro confermano le ipotesi iniziali riguardo l'efficacia dell'uso delle reti neurali artificiali per l'esecuzione musicale. Tali risultati inducono un ulteriore sviluppo del sistema realizzato. È importante notare come, con le reti ottenute con la presente ricerca, è possibile eseguire una qualunque partitura in tempo reale, conoscendo solo alcuni parametri fondamentali, che sono la durata nominale della nota, la sua intensità nominale e la sua altezza nel rigo musicale. Questo procedimento è una novità rispetto ai classici approcci per l'esecuzione mediante computer in cui o si vanno a tarare "a mano" i parametri esecutivi delle singole note oppure si usa un sistema esperto, che mediante regole simboliche appronta le caratteristiche espressive di ogni singola nota. Le differenze sostanziali dell'esecuzione mediante reti neurali rispetto ai due metodi citati sono due: la prima consiste nel fatto che con le reti neurali, una volta noti i valori di durata, intensità e altezza di ogni nota, l'esecuzione avviene in tempo reale, la seconda, e più importante, è che l'addestramento di una rete non dipende dal contesto, cioè si possono eseguire indifferentemente, ad esempio, musiche di Beethoven o di Mozart.

Un risultato importante è che, dopo un breve addestramento della rete (circa 15 minuti con un 80386 a 33 MHz) con uno stile particolare, è possibile far eseguire qualsiasi partitura con quello stesso stile ed in tempo reale: man mano che lo spartito viene "letto" dai neuroni di input, esso viene eseguito istantaneamente.

Un'altra conclusione interessante è che si è verificato il previsto "smussamento" delle deviazioni in output rispetto a quelle, esagerate, ottenute tramite l'effetto additivo del sistema di regole simboliche di Sundberg.

L'applicazione più importante di questo lavoro non è tanto l'esecuzione di partiture concepite per strumenti tradizionali, ma di musiche generate da computer, in quanto, per il momento, il computer non è in grado di effettuare autonomamente quelle microvariazioni, di tempo, intensità e timbro, che invece sono le caratteristiche fondamentali che contraddistinguono l'abilità esecutiva di un concertista. Una possibile applicazione del presente lavoro consiste nell'eseguire automaticamente delle partiture generate da computer, introducendo delle microvariazioni sia in

durata che in intensità, rendendo l'esecuzione più "armoniosa" all'orecchio umano.

In particolare sono state prese in considerazione le reti neurali discusse in precedenza che permettono di variare in tempo reale il parametro K delle regole di Sundberg e applicate all'uscita di una sorgente markoviana di melodie (Bresin, Manduchi 1989), che in pratica è un'improvvisazione automatica su un tema memorizzato in precedenza. Variando allora il parametro K durante l'esecuzione, potremo passare in qualunque istante da un'esecuzione "fredda" (bassi valori di K) ad una "romantica" (alti valori di K) a seconda dei gusti dell'operatore.

I risultati ottenuti possono anche avere un'applicazione commerciale, se pensiamo alla possibilità di trasferire le reti nelle ROM dei sintetizzatori, offrendo così la possibilità di esecuzioni automatiche "interpretate".

Con l'uso delle reti neurali si apre quindi una nuova strada nell'esecuzione musicale mediante computer.

RIFERIMENTI

- R. Bresin, R. Manduchi "Una sorgente di melodie con controllo di entropia", **Atti VIII Colloquio di Informatica Musicale**, Cagliari 26-28 ottobre 1989, pp. 213-215
- G. De Poli, L. Irone, A. Vidolin "Music score interpretation using a multilevel knowledge base", **Interface**, vol. 19, 1990, pp. 137-146
- A. Friberg "Generative Rules for Music Performance: A Formal Description of a Rule System", **Computer Music Journal**, vol. 15, No. 2, Summer 1991, pp. 56- 71
- A. Gabrielsson "Interplay Between Analysis and Synthesis in Studies of Music Performance and Music Experience", **Music Perception**, vol. 3, No. 1, 1985, pp. 59-86
- D.H. Hubel, "Il cervello", **Le Scienze**, No. 135, Novembre 1979
- D. Parisi, F. Cecconi, S. Nolfi "ECONETS: neural networks that learn in an environment", **Networks**, 1, 1990, pp. 149-168
- B.H. Repp "Patterns of expressive timing in performances of Beethoven minuet by nineteen famous pianists", **J. Acoust. Soc. Am.**, vol. 88, No. 2, August 1990, pp. 622-641
- B.H. Repp "Further Perceptual Evaluations of Pulse Microstructure in Computer Performances of Classical Piano Music", **Music Perception**, vol. 8, No. 1, 1990, pp. 1-33
- D.E. Rumelhart, J.L. McClelland **Parallel Distributed Processing**, vol. 1, Cambridge: MIT Press, 1988, chapter 8
- D.E. Rumelhart, J.L. McClelland **Explorations in Parallel Distributed Processing**, Cambridge: MIT Press, 1988b, chapter 5, appendix C
- J. Sundberg, A. Friberg, L. Frydén "Rules for automated performance of ensemble music", **Contemporary Music Review**, Vol. 3, 1989, pp. 89-109
- J. Sundberg et al. "Performance Rules for Computer-Controlled Contemporary Keyboard Music", **Computer Music Journal**, vol. 15, No. 2, Summer 1991, pp. 49-55
- J. Sundberg, W.F. Thompson, A. Friberg, L. Frydén "Use of rules for expression in the performance of melodies", **Psychology of music**, vol. 17, No. 1, 1989, pp. 63-82