

LA TRASCRIZIONE MUSICALE AUTOMATICA: CONFRONTO FRA METODI DI CLASSIFICAZIONE CON MEMORIA

Mario Salerno¹, Giovanni Costantini¹, Massimiliano Todisco¹, Renzo Perfetti²

¹Dipartimento di Ingegneria Elettronica
Università di Roma “Tor Vergata”
Via del Politecnico, 1 - 00133 Roma

²Dipartimento di Ingegneria Elettronica e dell’Informazione
Università di Perugia
Via Duranti, 93 - 06125 Perugia

La Trascrizione Musicale può essere definita come l’atto di ascoltare un brano musicale e trascriverlo secondo una notazione codificata che descriva le note che lo costituiscono. Quello della trascrizione musicale non è un compito facile: per persone senza un’educazione musicale è piuttosto difficile anche solo ripetere una frase musicale in maniera esatta dopo averla sentita, e la sua trasformazione in simboli non può in alcun modo avere luogo senza una familiarità sia, con la notazione stessa, sia con molte altre abilità musicali, quali la conoscenza dei meccanismi di produzione dei suoni, della teoria musicale e degli stili musicali, e non ultima l’esperienza e l’abitudine all’ascolto. La Trascrizione Musicale Automatica è il processo mediante il quale la Trascrizione Musicale è delegata a un algoritmo [1, 2, 3, 4].

Il nostro approccio utilizza strumenti di analisi la Trasformata a Q Costante (CQT) [5] e le Support Vector Machines [6]. Il riconoscimento delle note eseguite in un brano musicale è stato realizzato tramite classificatori di tipo SVM in architettura multiclasse OVA. Nelle sperimentazioni abbiamo utilizzato due diversi Kernel per le SVM: lineare e radiale (RBF). L’analisi spettrale utilizzata per le feature extraction è basata sulla CQT, inoltre, sono state utilizzate due diverse scale, lineare e logaritmica, per rappresentare le ampiezze della CQT. Si è focalizzata l’attenzione su due tipi di feature extraction: quelle con “memoria” e quelle senza “memoria”. Per “memoria” si intendono vettori che contengono anche i valori di CQT-bin della nota o delle note precedenti a quelle dell’evento nota attuale. In particolare, le scelte operative che riguardano il pre-processing con memoria sono:

- A - SENZA MEMORIA. Si effettua il calcolo di una sola CQT per onset trovato.
- B - MEMORIA BREVE. Si calcolano due CQT per ogni onset trovato: a partire dall’onset, una CQT è calcolata come nel caso senza memoria, l’altra è invece calcolata nell’istante immediatamente precedente all’onset.
- C - UN ONSET PRECEDENTE. Si calcolano due CQT per ogni onset trovato: a partire dall’onset, una CQT è calcolata come nel caso senza memoria, l’altra è invece la CQT relativa all’onset precedente.

Il sistema di Trascrizione completo è mostrato in Fig. 1. In particolare, il segnale musicale che attraversa il sistema di trascrizione, subisce due trattamenti:

- 1) una prima conversione operata dal sistema di pre-processing, che si occupa di analizzare il segnale cercando di rilevare i Midi onsets (onset detection) attraverso la STFT e isolare le armoniche dei suoni presenti in un determinato istante di tempo e che nel nostro caso è rappresentata dall’operazione di trasformazione a Q costante (CQT).
- 2) un processo di elaborazione, che a partire dalle componenti armoniche cerca di identificare il pitch di ogni nota presente in quell’istante di tempo, e che viene

realizzato mediante una classificazione supervisionata ad opera di una Support Vector Machine per ogni nota.

In Tabella I si riportano i risultati dell'accuratezza per ogni tipo di sistema. L'accuratezza maggiore è 85.7%, essa è stata ricavata con pre-processing a un onset precedente, Kernel radiale e ampiezze logaritmiche. Il successo di sistemi con memoria è dovuto al metodo di composizione dei brani che costituiscono il Dataset, basato sull'armonia classica, dove ogni nota musicale è strettamente correlata alla nota che la precede e che la segue nella composizione.

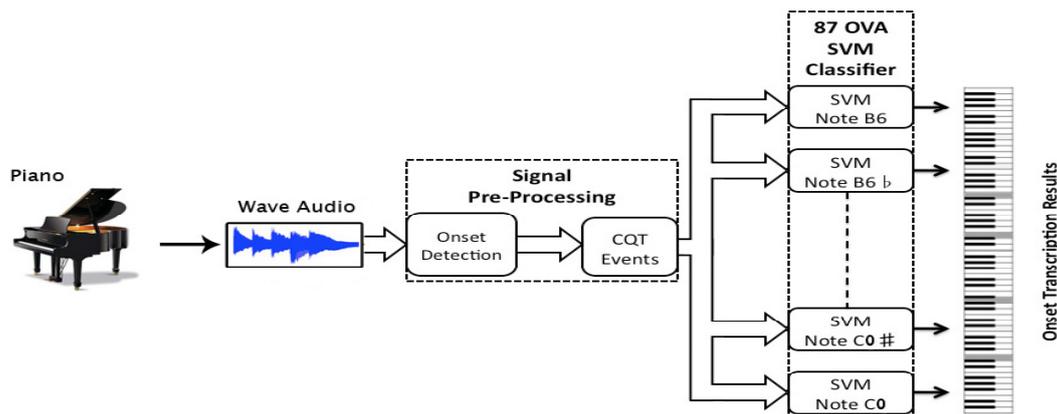


Fig. 1: Schema a blocchi del sistema di Trascrizione Musicale Automatica

SISTEMA	AMPIEZZA	KERNEL	
		LINEARE	RADIALE
SENZA MEMORIA	LINEARE	61.4%	66.6%
	LOGARITMICA	70.1%	74.9%
MEMORIA BREVE	LINEARE	73.4%	74.6%
	LOGARITMICA	76.7%	82.3%
UN ONSET PRECEDENTE	LOGARITMICA	75.5%	85.7%

Tabella I

Referenze

- [1] M. Marolt, "SONIC: A System for Transcription of Piano Music". Workshop on Current Research Directions in Computer Music 2001, Barcelona, Spain
- [2] G. Poliner and D. Ellis, "A Discriminative Model for Polyphonic Piano Transcription", EURASIP Journal of Advances in Signal Processing, vol. 2007, 48317, pp. 1-9, 2007.
- [3] G. Costantini, R. Perfetti, M. Todisco, (2009). Event based transcription system for polyphonic piano music. Elsevier, Signal Processing, vol. 89 (9); p. 1798-1811.
- [4] G. Costantini, M. Todisco, R. Perfetti, R. Basili, D. Casali: SVM Based Transcription System with Short-Term Memory Oriented to Polyphonic Piano Music, 15th MELECON IEEE Mediterranean Electrotechnical Conference, Valletta, Malta, April 26-28, 2010, p. 196-201.
- [5] Brown J. C. (1988), Calculation of a Constant Q Spectral Transform, J. Acoust. Soc. Am., 1991.
- [6] N. Cristianini, J. Shawe Taylor, "An introduction to Support Vector Machines (and other kernel-based learning methods)", Cambridge University Press, 2000.