

FEATURE EXTRACTION BASATO SU NMF PER SISTEMI DI TRASCRIZIONE MUSICALE AUTOMATICA

Giovanni Costantini^{1,2}, Massimiliano Todisco¹, Renzo Perfetti³

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

² Istituto di Acustica e Sensoristica “O.M. Corbino” - IDASC del CNR
Via del Fosso del Cavaliere, 100 - 00133 Roma,

³ Dipartimento di Ingegneria Elettronica, Università di Perugia
Via G. Duranti, 93 - 06125 Perugia

La Trascrizione Musicale è definita come l’atto di ascoltare un brano musicale e trascriverlo secondo una notazione codificata che descriva le note che lo costituiscono. Il compito è arduo: 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 della teoria musicale e degli stili musicali, dei meccanismi di produzione dei suoni, nonché 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 - 4].

L’approccio proposto utilizza strumenti di analisi quali la Trasformata a Q Costante (CQT) [5], le Orthogonal Non-negative Matrix Factorization (ONMF) [6] e le Support Vector Machines [7]. 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 e sulle ONMF. In particolare, data la matrice non negativa $M \times N$ delle CQT, $V \in \mathbb{R}^{\geq 0, M \times N}$, l’obiettivo è di approssimare V come il prodotto di due matrici non negative $W \in \mathbb{R}^{\geq 0, M \times R}$ e $H \in \mathbb{R}^{\geq 0, R \times N}$

$$V \approx WH \quad \text{s.t.} \quad H = W^T V \quad \text{e} \quad W^T W = I$$

dove R rappresenta il rango della fattorizzazione e richiede una conoscenza pregressa.

L’ONMF è un problema di ottimizzazione che ha una soluzione semplice se la funzione costo utilizzata è basata sulla norma di Frobenius $\min_{W \geq 0, H \geq 0} \|V - WH\|_F^2$, allora le regole iterative [6] di aggiornamento per W sono date da $W'_{ik} = W_{ik} \left(\frac{V V^T W}{W W^T V V^T W} \right)_{ik}$.

Il metodo proposto per l’estrazione delle feature è mostrato schematicamente nella Fig. 1.

La parte inferiore della figura rappresenta il segnale audio polifonico decomposto attraverso basi rappresentative previste a priori come modelli. Questi modelli sono appresi offline e costituiscono il dizionario utilizzato durante la decomposizione.

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 rilevando i Midi onsets (onset detection) attraverso la STFT e isolando le armoniche dei suoni presenti in un determinato istante di tempo, rappresentata dall’operazione di trasformazione a Q costante (CQT);

- 2) un modulo di decomposizione del segnale basato su ONMF;
- 3) un processo di elaborazione che, a partire dalle componenti armoniche, identifica il pitch di ogni nota presente in quell'istante di tempo, realizzato mediante una classificazione supervisionata ad opera di una Support Vector Machine per ogni nota.

Il sistema, ancora in fase di test su tutte le 84 note, ha dato dei miglioramenti in accuratezza del 4% circa su note del registro medio rispetto ad analoghi sistemi che non usano tecniche NMF presentati in [3].

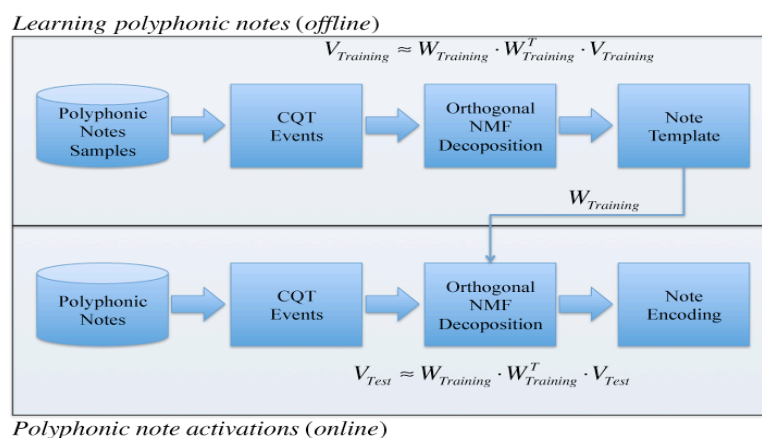


FIGURA 1: Sistema di Trascrizione Musicale Automatico

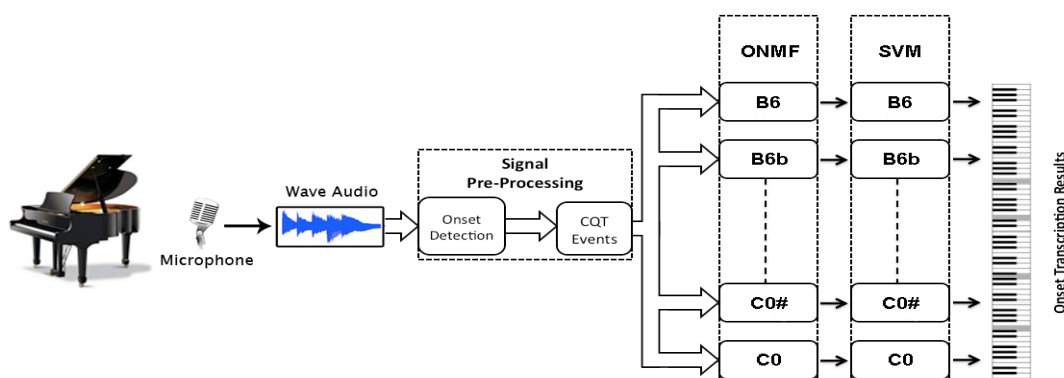


FIGURA 2: Sistema di Trascrizione Musicale Automatico

Riferimenti

- [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] J. C. Brown, Calculation of a Constant Q Spectral Transform, J. Acoust. Soc. Am., 1991.
- [6] Z. Yang and E. Oja, Linear and Nonlinear Projective Nonnegative Matrix Factorization, IEEE Transactions on Neural Networks, 21(5):734-749, 2010.
- [7] N. Cristianini, J. Shawe Taylor, "An introduction to Support Vector Machines (and other kernel-based learning methods)", Cambridge University Press, 2000.