

ALGORITMO DI PARAMETRIZZAZIONE DI MODELLI FISICI ACUSTICI BASATO SU APPRENDIMENTO END-TO-END

L. Gabrielli, S. Tomassetti, P. Vecchiotti, S. Squartini, F. Piazza

Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche
Via Brecce Bianche 12, 60131, Ancona, Italia, e-mail s.squartini@univpm.it

Nell'ambito dell'acustica computazionale, l'impiego di modelli fisici ha permesso di raggiungere importanti traguardi nella generazione numerica del suono di strumenti musicali, con la capacità di rendere la qualità timbrica molto veritiera. Tuttavia più si cerca una rappresentazione fedele del fenomeno fisico più il numero di parametri del modello digitale cresce, rendendo l'operazione di stima di tali parametri complessa, specie in modelli più astratti, che adottano delle assunzioni semplificative al fine di ridurre il costo computazionale come nei modelli di propagazione con Digital Waveguide (DWG) [1].

In passato, la parametrizzazione di modelli fisici è stata approcciata in due differenti modi: il primo è un metodo white-box, come in [2] dove gli autori mappano direttamente delle costanti fisiche all'interno dei parametri del modello misurandole dopo aver computato la DFT. Per questo tipo di approccio è necessaria una modellazione molto precisa ed è indispensabile un collegamento tra i parametri del modello e le grandezze fisiche. Il secondo metodo è un metodo black-box col quale si opera direttamente nel dominio del modello fisico. In [3] gli autori sfruttano una Least-Square Optimization per la stima di parametri di un modello di clarinetto ad ancia. Gli autori di [4] e [5] utilizzano un algoritmo genetico, quindi una ricerca iterativa dei parametri. Storicamente uno dei primi approcci di questo genere utilizza una rete neurale Multi-Layer Perceptron (MLP) per parametrizzare un semplice modello fisico a DWG [6].

Il modello fisico utilizzato in questa ricerca come caso applicativo è un Flue-Pipe Organ Model [7] realizzato e commercializzato dall'azienda Viscount International Spa, sviluppato con il paradigma DWG e controllato da 58 parametri di cui 32 di essi sono relativi ad involucri e ampiezze, i restanti vanno a modellare la parte armonica del suono attraverso filtri, delay lines e coefficienti lineari e non lineari. I parametri sono strettamente interconnessi e interdipendenti; inoltre la generazione di un determinato suono non è univoca e quindi è ottenibile sfruttando svariate differenti configurazioni di parametri. Il sistema di apprendimento end-to-end utilizzato è basato su Convolutional Neural Networks (CNN) che a differenza di [6] non necessitano di features customizzate, ma semplicemente del segnale *raw* nel dominio del tempo, o una sua rappresentazione tempo-frequenza. Il dataset di training è composto da un set di suoni, denominati *contrived*, provenienti dalla libreria del modello stesso e precedentemente elaborati. Di questo se ne sono variate dimensioni e composizione negli esperimenti. Il dataset di test è invece composto da suoni *contrived* non inclusi nel training set e *non-contrived*, cioè suoni provenienti dal mondo reale, quindi campioni da riprodurre quanto più fedelmente possibile sfruttando il modello fisico. Un differente dataset utilizzato in training è quello denominato *random*, cioè un dataset generato campionando uniformemente lo spazio dei parametri del modello. Uno studio preliminare sul modello di Karplus-Strong [7] ha portato ad un confronto tra metodi di campionamento dello spazio dei parametri e conseguente scelta del più performante per i training. Il metodo end-to-end proposto è stato testato con la STFT in modulo e fase del segnale d'ingresso, la STFT solo modulo e i coefficienti Log-Mel, molto utilizzati in ambito audio come ingresso per reti convoluzionali. La misura delle prestazioni è data in termini di Mean Absolute Error (MAE) sui parametri del modello, calcolabile solo su

testing se *contrived*, Log Spectral Distance (LSD) e Harmonic Mean Squared Error sulle prime 10 armoniche (HMSE10). I risultati sono riportati in Tabella 1:

Training set	Test set	Input	MAE	LSD (dB)	HMSE10 (dB)
Contrived	Contrived	STFT mag+phase	0.205	11.8	6.6
		STFT power spectrogram	0.25	13.0	13.1
		MEL spectrogram	0.252	18.7	13.3
Random	Contrived	STFT mag+phase	0.432	76.3	19.2
		STFT power spectrogram	0.471	89.0	23.3
		MEL spectrogram	0.446	83.4	23.9
Contrived	Non-contrived	STFT mag+phase	N/A	30.1	15.6
		STFT power spectrogram	N/A	32.8	16.3
		MEL spectrogram	N/A	35.4	17.9

Tabella 1: Prestazioni dell’algoritmo di parametrizzazione End-to-End in termini di MAE, LSD e HMSE10.

Per quanto riguarda la possibilità di sviluppi futuri si avrà modo di raffrontare l’approccio proposto con altri non del tipo End-to-End quindi valutando una estrazione mirata delle features. Un’ulteriore possibilità è data dall’impiego diretto del modello fisico nella ricerca dei parametri usando un algoritmo di apprendimento o un algoritmo di ricerca iterativa. I risultati di questa ricerca sono stati presentati su due pubblicazioni internazionali [8,9].

BIBLIOGRAFIA

- [1] V. Välimäki, J. Pakarinen, C. Erkut, and M. Karjalainen, “Discrete-time modelling of musical instruments,” *Reports on Progress in Physics*, vol. 69, pp. 1–78, 2006.
- [2] V. Chatziioannou and M. van Walstijn, “Estimation of clarinet reed parameters by inverse modelling,” *Acta Acustica united with Acustica*, vol. 98, no. 4, pp. 629–639, 2012.
- [3] J. Riionheimo and V. Välimäki, “Parameter estimation of a plucked string synthesis model using a genetic algorithm with perceptual fitness calculation,” *EURASIP Journal on Advances in Signal Processing*, vol. 2003, no. 8, 2003.
- [4] C. Drioli and D. Rocchesso, “A generalized musical-tone generator with application to sound compression and synthesis,” in *Acoustics, Speech, and Signal Processing, 1997 IEEE International Conference on*. IEEE, 1997, vol. 1, pp. 431–434.
- [5] A.T. Cemgil and C. Erkut, “Calibration of physical models using artificial neural networks with application to plucked string instruments,” *Proc. Intl. Symposium on Musical Acoustics (ISMA)*, vol. 19, pp. 213–218, 1997.
- [6] C. Zinato, “Method and electronic device used to synthesize the sound of church organ flue pipes by taking advantage of the physical modeling technique of acoustic instruments” Oct. 28 2008, US Patent 7,442,869.
- [7] K. Karplus and A. Strong, “Digital synthesis of plucked-string and drum timbres,” *Computer Music Journal* 7(2): 43-55, 1983.
- [8] L. Gabrielli, S. Tomassetti, C. Zinato, and S. Squartini, “Introducing deep machine learning for parameter estimation in physical modelling”, *Proc. Digit. Audio Effects – DAFX2017*, 11-16, 2017.
- [9] L. Gabrielli, S. Tomassetti, C. Zinato, F. Piazza, “End-to-End Learning for Physics-based Acoustic Modeling”, *IEEE Trans. Emerging Topics in Computational Intelligence*, Apr.2018.