POLITECNICO
MILANO 1863

Cremona, 8 giugno 2021

POLO TERRITORIALE DI CREMONA

SESSIONE DI LAUREA MAGISTRALE AL CAMPUS DI CREMONA DEL POLITECNICO DI MILANO

ELENCO DEI LAUREANDI CON TITOLO TESI E ABSTRACT:

De LUCA GIORGIO

Titolo tesi: Modeling perceived rhythmic complexity using multiple complexity measures

Abstract

Stabilire la complessità di un ritmo percepita da un ascoltatore in maniera deterministica è un argomento di particolare interesse nel campo dell'ingegneria musicale, specialmente se si considera l'uso sempre più ampio di applicazioni basate su Music Information Retrivial, come la classificazione e i suggerimenti di musica affine. Sebbene la complessità ritmica sia una caratteristica fortemente soggettiva di un ritmo, in letteratura ci sono diversi metodi che permettono di stimarla. La maggior parte di queste misure fanno riferimento solo a singoli aspetti coinvolti nella complessità, mentre la mente umana considera simultaneamente più fattori. Dunque la complessità ritmica non può essere unicamente caratterizzata dall'uso di una singola metrica, ma richiede una loro combinazione al fine di ottenere un modello della percezione della complessità ritmica più robusto. In questa tesi proponiamo un nuovo approccio per modellare la percezione della complessità ritmica basato sull'utilizzo di più misure di complessità presenti in letteratura. Con lo scopo di identificare l'insieme di misure più rilevanti e meno ridondanti, abbiamo studiato le prestazioni di ben noti algoritmi di feature selection e discusso successivamente i loro limiti. Infine, presentiamo un nuovo algoritmo capace di generare un ranking delle misure di complessità. Questo ranking può essere effettivamente usato per ridurre la complessità del modello. Per concludere, il metodo da noi proposto è in grado di offrire modelli dall'alto potere espressivo al tempo stesso conservativi rispetto alla percezione della complessità ritmica.

EPIFANI EDOARDO

Titolo tesi: Perception of ITDG, Reverberation time, and density of impulse response of a room for a different reverberation time

Abstract

Questo studio di psicoacustica esplora come vengono percepiti tre parametri di un riverbero, l'Initial Time Delay Gap (ITDG), Reverberation Time (RT) e la densità delle prime riflessioni usando due tipi di campioni sonori e riverberi con lunghezze differenti per i primi due parametri citati. Test creati appositamente su di un sito web con metodi adattativi usati per

Polo Territoriale di Cremona
Via Sesto, 41
26100 Cremona

Tel. 0372 56 7711
Fax 0372 56 7701
segreteria-cremona@polimi.it
www.polo-cremona.polimi.it

Partita Iva 04376620151
Codice Fiscale 80057930150

POLITECNICO
MILANO 1863

selezionare il tipo di suono adatto e metodi alternative forced-choice per fare in modo che il tester fosse nella migliore condizione per rispondere in base al tipo di test proposto. L'analisi proposta mostra una forte correlazione tra la percezione dell'RT e la durata di un riverbero, cosa che non si può affermare per l'ITDG perché poco correlato. Per quanto riguarda il tipo di sample utilizzato, l'ITDG è lievemente influenzato rilevando così che un campione vocale ha ottenuto in media risultati minori, invece l'esito del test sulla densità dimostra che viene distinto più facilmente un riverbero con densità delle prime riflessioni più alta usando un campione strumentale, i risultati dei test dell'RT non rilevano alcuna differenza se non trascurabile. Infine si può sostenere che i risultati ottenuti sono simili a quelli ottenuti in altri studi con un margine di errore maggiore causato dal fatto di non esser stati svolti tutti nelle condizioni ottimali e da persone esperte di ascolto musicale.

MORI DAVIDE

Titolo tesi: Towards soundfield rendering with distributed loudspeaker arrays using Convolutional Neural Networks applied to the Ray Space Transform

Abstract

Il problema della riproduzione di campi acustici è stato di fondamentale interesse di ricerca per anni, grazie ai suoi vari campi applicativi. L'obiettivo della tesi è la riproduzione, all'interno di un'area d' ascolto, del campo acustico generato da sorgenti acustiche virtuali attraverso un setup costituito da una schiera lineare uniforme di altoparlanti con alcuni di essi mancanti, risultandone in uno spaziamento irregolare. Per affrontare il problema, vogliamo derivare i coefficienti di riproduzione degli altoparlanti usando una rappresentazione intermedia del campo acustico, quale la Trasformata Ray Space RST. Essa mappa le informazioni acquisite da una schiera di microfoni precedentemente posizionato nella stessa posizione della schiera di altoparlanti, nel dominio dello spazio dei raggi: qui una sorgente puntiforme corrisponde a un raggio nello spazio euclideo, ossia una linea orientata sulla quale viene trasportata l'energia acustica. Inoltre, scegliamo come guida un metodo di riproduzione basato sulla decomposizione in onde piane PWD con tecniche di filtraggio spaziale per calcolare i coefficienti di riproduzione e i campi risultanti. In questo lavoro, utilizziamo un approccio basato sulle Reti Neurali Convoluzionali su un problema di regressione. Abbiamo diviso il problema in due fasi: nella prima, la rete deve imparare una corrispondenza tra l'immagine della RST in ingresso e i coefficienti di riproduzione per ricostruire il campo acustico nel caso di una schiera completa. Nella seconda, a causa della mancanza di altoparlanti e dei microfoni precedentemente posizionati, viene utilizzata la versione degradata della RST; si sfrutta quindi il Transfer Learning per compensare l'assenza di tali altoparlanti per una riproduzione accurata

del campo finale. Per dimostrarne l'efficacia, mostriamo le condizioni in cui il nostro metodo supera quello basato su PWD, contando di minimizzare l'errore rispetto al campo riprodotto da una data sorgente reale nella stanza. I risultati di questo lavoro possono essere visti come un primo passo verso un problema con schiere di altoparlanti distribuiti, utilizzando il Deep Learning combinato con la Trasformata Ray Space.

PANTALEONE AGNESE

Titolo tesi: Modeling Multichannel Room Impulse Responses using Banks of State-Space Filters

Abstract

Negli ultimi decenni, il riverbero artificiale ha acquisito un ruolo principale nel campo della produzione e riproduzione audio. La comunità di ricerca ha proposto nel tempo approcci anche molto diversi tra loro, ma tutti comunque risentono dell'incompatibilità che caratterizza la necessità di raggiungere risultati altamente accurati e allo stesso tempo il bisogno di limitare la complessità computazionale. In particolare, nel caso dei sistemi multi-canale, questo trade-off tra le due esigenze è molto più difficile da raggiungere. Il metodo proposto, il cui fine è quello di indagare ulteriormente nella ricerca di un compromesso ottimale tra accuratezza e semplicità di calcolo, può essere contestualizzato nel mezzo tra le tecniche basate sull'utilizzo della convoluzione e i metodi approssimati che sfruttano sistemi composti da delay networks. Per raggiungere l'obiettivo di descrivere fedelmente in termini matriciali una Room Impulse Response (RIR) di riferimento, la procedura presentata in questa tesi fa uso dell'Eigensystem Realization Algorithm, un algoritmo ampiamente utilizzato per il controllo dei sistemi data-driven, la cui conoscenza si basa esclusivamente su un ampio set di dati provenienti da misurazioni e/o simulazioni. La procedura prevede una prima fase in cui la RIR di riferimento viene suddivisa in blocchi, i quali a loro volta possono essere trattati in maniera indipendente l'uno dall'altro applicando diversi livelli di riduzione della dimensione del modello. In questo modo, la RIR in esame viene rappresentata da un banco di filtri state-space descritti tramite matrici. Di conseguenza, questi filtri vengono poi utilizzati per processare un generico segnale di ingresso al quale dev'essere aggiunto del riverbero. Questo metodo di implementazione a blocchi può essere applicato in entrambi i casi SISO e MIMO e gode di un alto grado di flessibilità. Infatti, in base al numero di segmenti scelto e al livello di riduzione dell'ordine applicato, permette di adattare la propria complessità computazionale allo specifico scenario di riverbero artificiale considerato. Infine, come sviluppo futuro di questo lavoro di ricerca, riteniamo che il metodo proposto possa aprire la strada verso riverberatori artificiali più avanzati basati su implementazioni a blocchi, ed allo stesso tempo più efficienti dal punto di vista computazionale, in quanto caratterizzati sia dalla presenza di filtri state-space per la

descrizione delle prime riflessioni, che da FDN, utilizzati per modellare invece la coda di riverbero.

PINO FRANCESCO

Titolo tesi: An approach to soundfield rendering and loudspeaker placement in free-field conditions

Abstract

Questa tesi propone un nuovo metodo di design dei filtri per la riproduzione del campo sonoro e un approccio per l'ottimizzazione delle posizioni dei loudspeakers basato sul metodo di riproduzione proposto, in un ambiente senza riverbero. Il campo sonoro generato dalla sorgente virtuale è approssimato elaborando con filtri spazio-temporali i segnali emessi dagli altoparlanti. Tali filtri spazio-temporali sono progettati risolvendo un problema di minimizzazione dei minimi quadrati. L'ottimizzazione della posizione degli altoparlanti è ottenuta applicando un algoritmo iterativo per la minimizzazione di una funzione di costo opportunamente definita. In particolare, data una condizione iniziale, cioè un array di altoparlanti con posizioni non ottimizzate, l'algoritmo restituisce le posizioni degli altoparlanti che minimizzano la funzione di costo definita. In conclusione, vengono descritti due scenari di applicazione dell'approccio proposto: nel primo scenario l'ottimizzazione della geometria viene eseguita dopo aver fissato il numero di altoparlanti disponibili; nel secondo scenario l'ottimizzazione viene eseguita utilizzando un numero illimitato di altoparlanti dopo aver impostato un errore di ricostruzione del campo sonoro. L'analisi di entrambi gli scenari viene eseguita considerando tre tipi di geometrie di array 2D da ottimizzare: array lineari uniformi, array lineari non uniformi e array ad arco.

SCERBO MATTEO

Titolo tesi: Speech separation and diarization with microphone arrays: integrating tasks to improve performance

Abstract

I campi di Diarization --- determinare chi ha parlato quando, in una conversazione --- e Speech Separation --- isolare una voce da rumori e altre voci --- hanno visto grandi progressi negli anni recenti. In particolare, molte pubblicazioni hanno proposto nuovi modi di sfruttare gli aspetti condivisi di queste applicazioni, avanzando entrambi i campi di ricerca. Lo scopo di questa tesi è realizzare una completa integrazione di tali campi. Per questo scopo, abbiamo progettato e allenato una singola rete neurale che si occupi di ambo i compiti, estraendo dal segnale sonoro caratteristiche del parlato utili per l'identificazione come per l'estrazione. Il nostro sistema, che effettua simultaneamente separazione e identificazione vocale, è composto da tre moduli indipendenti. Per prima cosa, tramite un array di microfoni, viene stimata la direzione d'arrivo

delle voci di ogni interlocutore attivo in un determinato istante. Viene quindi applicato un beamformer verso ognuno di essi, compiendo una separazione preliminare. I segnali dei beamformer sono successivamente processati da una rete neurale, la quale produce due output: un embedding che caratterizza l'identità dell'interlocutore, e una maschera di separazione da applicare al segnale del beamformer per isolarne la voce. Infine, viene eseguito un clustering degli embedding per associare ognuno alla relativa identità, e i segnali mascherati sono assegnati a canali di output diversi in base all'identificazione. La rete neurale è stata progettata tenendo a mente la condivisione delle feature tra i due compiti, e per questo è stata scelta un'architettura di tipo U-Net. Gli embedding vengono estratti dal collo di bottiglia, mentre le maschere di separazione sono l'output del decoder. Inoltre è stato prodotto un dataset di conversazioni simulate in stanze riverberanti, con il quale è stato allenato e testato il sistema. I risultati mostrano efficacia di identificazione eccellente e qualità di separazione significativamente superiore rispetto al segnale del beamformer.