

CONSERVATORIO "F. A. BOPNPORTI"

UNIVERSITÀ DEGLI STUDI DI TRENTO
DIPARTIMENTO DI INGEGNERIA E SCIENZA DELL'INFORMAZIONE

CORSO DI SISTEMI E LINGUAGGI DI PROGRAMMAZIONE PER L'AUDIO

Anno Accademico 2018/2019

Analisi dello Spettro e Riconoscimento di Modelli Noti

Approfondimento teorico e patch per Max 5.1

Nicola Fiorello

Docente: Mauro Graziani

Versione 1.0

Patch e documentazione disponibili online:

works.nicolafiorello.it/asrmn

Trento, 22/03/2019

Indice

1	Introduzione	2
1.1	Analisi su n componenti	2
1.2	Spettro e vocali	2
1.3	Sintesi additiva	3
1.4	Obiettivo	3
2	Struttura della Patch	4
2.1	Input	4
2.2	Filtro <i>biquad</i>	4
2.3	Analisi e raccolta dati	5
2.4	Sintetizzatore additivo	6
2.5	Array sperimentali	6
2.6	Valutazione dell'errore complessivo	7
2.7	Decisione per massima verosimiglianza	8
3	Conclusioni	9
3.1	Risultati ottenuti	9
3.2	Possibili migliorie e sviluppi	9

1 Introduzione

Lo spettro di un segnale è la sua descrizione nel dominio delle frequenze, spazio matematico in cui lo studio del segnale stesso è agevolato sotto molti aspetti.

Qualunque funzione periodica, continua e limitata può essere rappresentata mediante una somma di funzioni sinusoidali pure di opportuna ampiezza, frequenza e fase.

Teorema di Fourier

Analizzare lo spettro significa sostanzialmente osservare quanto impatta ogni componente in frequenza sul segnale complessivo.

I suoni "intonati", detti altresì "armonici", hanno la caratteristica di riportare, nello spettro delle frequenze, contributi disposti su multipli di un valore base, chiamato "frequenza fondamentale". Analizzare, in un suono, quanto importanti sono i contributi ad ogni multiplo intero della fondamentale permette ad esempio di apprezzarne la caratterizzazione di timbro.

1.1 Analisi su n componenti

Dal punto di vista teorico, un qualsiasi suono potrebbe richiedere, per essere rappresentato correttamente nel suo spettro, contributi su un infinito numero di frequenze. Questo scenario è chiaramente impossibile da realizzare o rappresentare con qualsiasi calcolatore, ma è possibile semplificare parecchio il modello isolando solo le prime n frequenze (nel caso armonico, i primi n multipli interi della fondamentale) e studiando l'intensità di questi contributi.

Una soluzione pratica per analizzare quindi lo spettro su un numero finito di frequenze è quello di filtrare per banda il segnale audio sulle n frequenze interessate, e di misurare il volume in uscita da ogni filtro. In questo modo è possibile rendersi conto di quanto, quella frequenza, sia influente sul suono complessivo. Filtrare un suono significa sostanzialmente attenuare tutto ciò che non rientra in un determinato intervallo di frequenze.



Figura 1: Spettro completo di un suono a confronto con lo stesso filtrato attorno ad un armonico.

1.2 Spettro e vocali

Come descritto precedentemente, l'analisi dello spettro permette di riconoscere un timbro rispetto ad un altro. Così come strumenti musicali diversi generano suoni con spettri strutturalmente differenti, anche le diverse vocali che la voce umana può cantare sono caratterizzate da un comportamento spettrale specifico.

Nel dettaglio, già isolando da un campione di voce umana i contributi della fondamentale e dei primi 5 armonici ($n = 6$), si può apprezzare un'importante differenziazione della distribuzione dell'intensità sulle diverse componenti. Da uno studio condotto sperimentalmente, si nota che le vocali "A" e "U" sono le due che presentano, tra loro, la più evidente differenziazione.

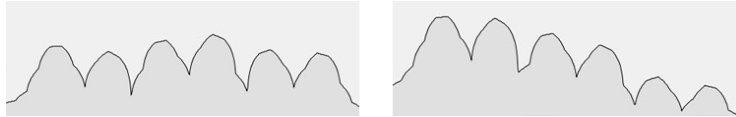


Figura 2: Spettro filtrato sulle prime 6 componenti delle vocali (rispettivamente) "A" e "U".

1.3 Sintesi additiva

Tra le tecniche di sintesi dell'audio, quella definita "Additiva" si caratterizza per la costruzione del suono tramite somma di sinusoidi pure. Per quanto detto prima, un qualsiasi suono è teoricamente riproducibile sommando sinusoidi aventi opportune caratteristiche.

Nel mondo dei suoni armonici, l'analisi dello spettro, svolta isolando le componenti tramite filtraggio e misurandone l'intensità, può fornire dati interessanti per pilotare gli oscillatori di un sintetizzatore additivo. Il suono rigenerato avrà una fedeltà al suono originale crescente al crescere del numero dei filtri impiegati (e quindi delle componenti valutate) e risentirà inoltre del posizionamento di tali filtri.

1.4 Obiettivo

L'obiettivo del progetto è creare un analizzatore di spettro che, isolate le prime 6 componenti di un suono, ne misuri l'intensità e valuti quanto queste influiscono sulla totalità del segnale. Una serie di oscillatori potranno poi essere pilotati con i valori raccolti in analisi, al fine di apprezzare la somiglianza di suoni rigenerati secondo ampiezze misurate da sorgenti reali.

Chiaramente, la qualità del suono riprodotto e l'efficacia del processo di analisi e ricostruzione saranno molto limitati a causa del basso numero di componenti che si è scelto di analizzare; nel caso delle vocali, ad esempio, sarà probabilmente possibile riconoscere quale delle due candidate è stata analizzata come modello per rigenerare il suono, ma il risultato sarà molto lontano da un suono naturale.

Il segnale in ingresso, analizzato secondo la tecnica sopra descritta, sarà poi confrontato con dei modelli raccolti sperimentalmente, e un semplice algoritmo permetterà al calcolatore di decidere per quale vocale si sta cantando.

2 Struttura della Patch

La patch realizzata è un unico file, costruito e testato su *Max 5.1*. È possibile utilizzare l'analizzatore su un file audio, caricando un file e riproducendolo, oppure su un input continuo da microfono. I valori raccolti continuamente possono essere poi "congelati" ed inviati al sintetizzatore per generare un suono il più simile possibile a quello dato in input. Un algoritmo di decisione per massima verosimiglianza deciderà inoltre quale delle vocali si sta cantando, e la mostrerà all'utente. È sempre possibile ascoltare il suono filtrato sulle componenti interessate.

2.1 Input

Come già detto, gli input possibili sono quello continuo tramite microfono (attraverso il componente *ezadc*) e quello tramite file, che sfrutta il componente *sfplay*.

I due input sono sommati assieme (il primo attraverso un regolatore di volume, il secondo direttamente) ed inviati all'analizzatore di spettro.

2.2 Filtro *biquad*

Il componente *biquad* è un filtro dinamico che può svolgere tagli di diverso tipo. Il taglio adatto per il nostro utilizzo è sicuramente il "passa-banda", che attenua tutto ciò che non rientra in una banda incentrata in una data frequenza.

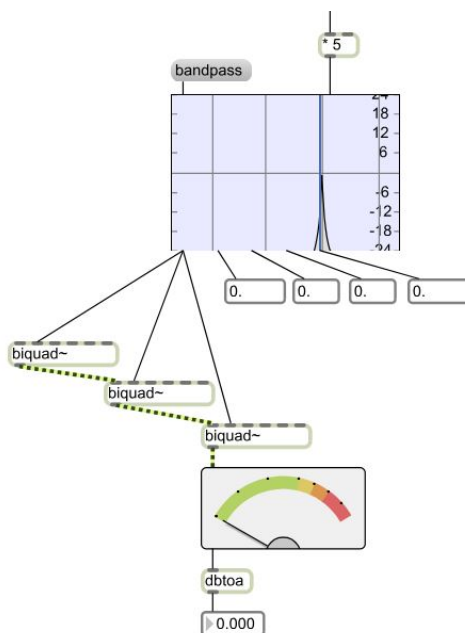


Figura 3: Componente filtrante, misura l'intensità del quinto armonico

Il filtro è pilotato dal componente *filtergraph*, che può essere impostato graficamente e fornisce a *biquad* tutti i dati di cui ha bisogno per eseguire il suo taglio, come frequenza

centrale, larghezza della banda, pendenza della linea di taglio, eventuale guadagno, ecc. L'utilizzo di 3 filtri in serie permette di effettuare un taglio il più stretto possibile, ed aumenta quindi l'efficienza del processo di selezione delle componenti. In uscita dall'ultimo filtro è stato collocato un *levelmeter*, che permette di valutare numericamente l'intensità del suono filtrato, e un *dbtoa*, che converte questo dato da una scala logaritmica a quella lineare.

2.3 Analisi e raccolta dati

Specificata una frequenza fondamentale di lavoro, i 6 filtri realizzati su modello di quello descritto al punto precedente portano le loro frequenze centrali automaticamente sulla fondamentale e sui primi 5 armonici. I 6 valori di volume raccolti vengono memorizzati in una serie di variabili, e i 6 segnali audio vengono sommati in un unico suono, che viene riprodotto in uscita tramite *ezdac* e mostrato graficamente all'utente tramite il componente *spectroscope*.

I diversi valori di volume relativi alle diverse componenti armoniche vengono poi normalizzati su una scala percentuale, in maniera da rendere più comprensibile l'importanza con cui ogni frequenza contribuisce al suono complessivo.

Il calcolo del contributo percentuale C avviene, per gli n oscillatori, con una semplice formula:

$$C_k = 100 \times \frac{V_k}{\sum_{i=1}^n [V_i]}$$

dove V_k è l'ampiezza volumetrica del k-esimo componente.

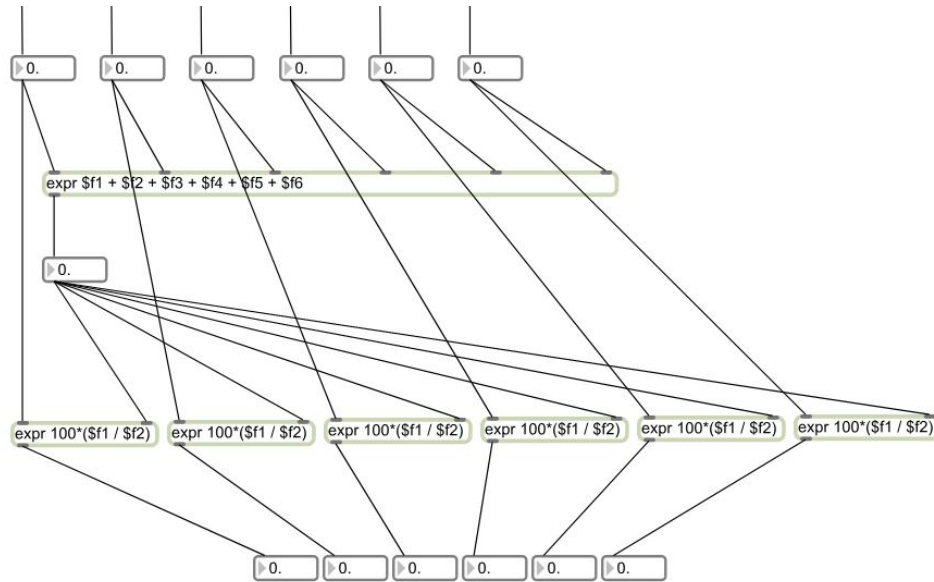


Figura 4: Calcolo del contributo percentuale.

Il set di dati raccolti permette, previa attivazione di un apposito *toggle*, di aggiornare i parametri che pilotano i 6 oscillatori. Un *bang* è necessario per effettuare istantaneamente la retroazione.

2.4 Sintetizzatore additivo

È stato realizzato un semplice sintetizzatore in additiva, che somma tra loro le sinusoidi generate da 6 oscillatori, rappresentati dai componenti *cycle*.

Impostata la frequenza fondamentale, le altre 5 vengono calcolate automaticamente al fine di generare sinusoidi armoniche, ossia con frequenze multiple della fondamentale. I valori di intensità delle singole sinusoidi possono essere "importati" dall'output dell'analizzatore di spettro, al fine di riprodurre un suono sintetizzato che si avvicini (per quanto reso possibile dal numero di oscillatori disponibili) al suono analizzato.

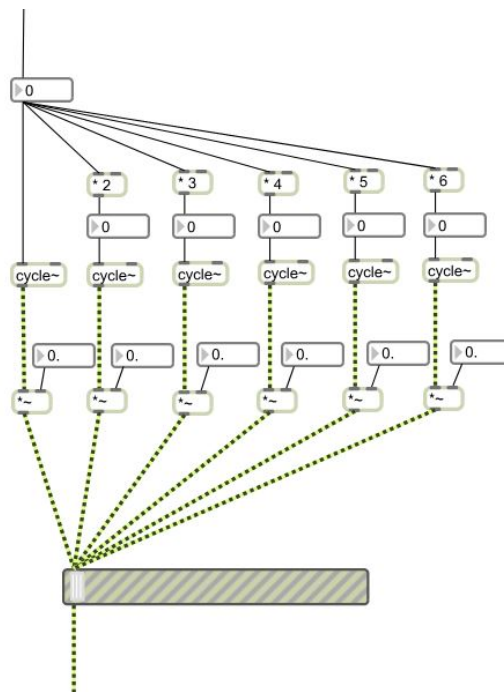


Figura 5: Struttura del sintetizzatore.

Il risultato della somma dei 6 segnali viene poi attenuato da un controllo di volume, e mescolato agli input della patch.

NOTA: Il controllo della frequenza fondamentale del sintetizzatore imposta anche le frequenze di lavoro dei filtri e degli altri componenti dell'analizzatore. È quindi indispensabile specificare la frequenza di lavoro indipendentemente dall'utilizzo del sintetizzatore.

2.5 Array sperimentali

Al fine di poter testare un algoritmo di decisione, sono stati realizzati sperimentalmente degli array contenenti i valori di intensità sui diversi armonici delle due vocali "A" e "U".

La realizzazione è avvenuta tramite misurazione di un suono continuo proveniente dal microfono mediante l'analizzatore di spettro realizzato; sono stati raccolti i dati relativi a

due istanti differenti, in cui sono state cantate le due vocali sulla frequenza fondamentale, normalizzati in contributo percentuale sulle sei componenti.

Per migliorare l'accuratezza della misura sono stati salvati gli array risultanti dalla media aritmetica degli array relativi a 4 diverse misurazioni della stessa vocale.

A					
16,20	10,22	22,89	45,67	6,50	8,14
17,00	9,56	19,08	30,24	15,32	9,66
13,73	7,72	21,76	30,74	11,53	10,28
15,44	8,68	27,45	30,80	9,43	5,30
15,59	9,05	22,80	34,36	10,70	8,35

Si nota che, nell'analisi della vocale "A", la tipica attenuazione graduale delle componenti è meno evidente, e vi sono addirittura misurazioni (sugli armonici terzo e quarto) che superano quella della fondamentale.

U					
47,85	38,01	15,13	4,27	0,50	0,16
40,49	36,09	12,80	5,09	0,38	0,24
44,28	31,35	11,12	6,26	0,53	0,21
44,80	35,58	10,03	3,99	0,37	0,29
44,36	35,26	12,27	4,90	0,45	0,23

La vocale "U", invece, ha un'attenuazione più evidente: nessun armonico è più impattante dei precedenti, e la fondamentale contribuisce per quasi metà dell'intensità sonora totale.

2.6 Valutazione dell'errore complessivo

Al fine di poter compiere una decisione, si è deciso di procedere con un algoritmo di massima verosimiglianza. I dati raccolti continuativamente sono stati confrontati con i dati raccolti dall'esperimento sulle vocali "A" e "U", e si è sviluppato un misuratore dell'errore complessivo commesso rispetto ai due modelli.

Si tratta sostanzialmente di un valore che indica la "distanza" tra il suono ricevuto in input e il modello memorizzato negli array sperimentali. Un valore alto è quindi associato alla disuguaglianza dei due suoni confrontati, mentre la somiglianza tra i due produrrà valori d'errore complessivo bassi.

L'errore E è stato calcolato con la seguente formulazione matematica:

$$E = \sum_{k=1}^n |C_k - C'_k|$$

dove C_k e C'_k sono rispettivamente il contributo percentuale della componente k -esima nel suono in input (analizzato continuativamente) e il contributo percentuale memorizzato nell'array sperimentale.

Sono stati realizzati due schemi di valutazione dell'errore, da far funzionare con i due diversi array di dati raccolti; il funzionamento finale, quindi, è quello di stimare quanto "distanti" ci si trova dalla vocale cantata "A" e dalla "U".

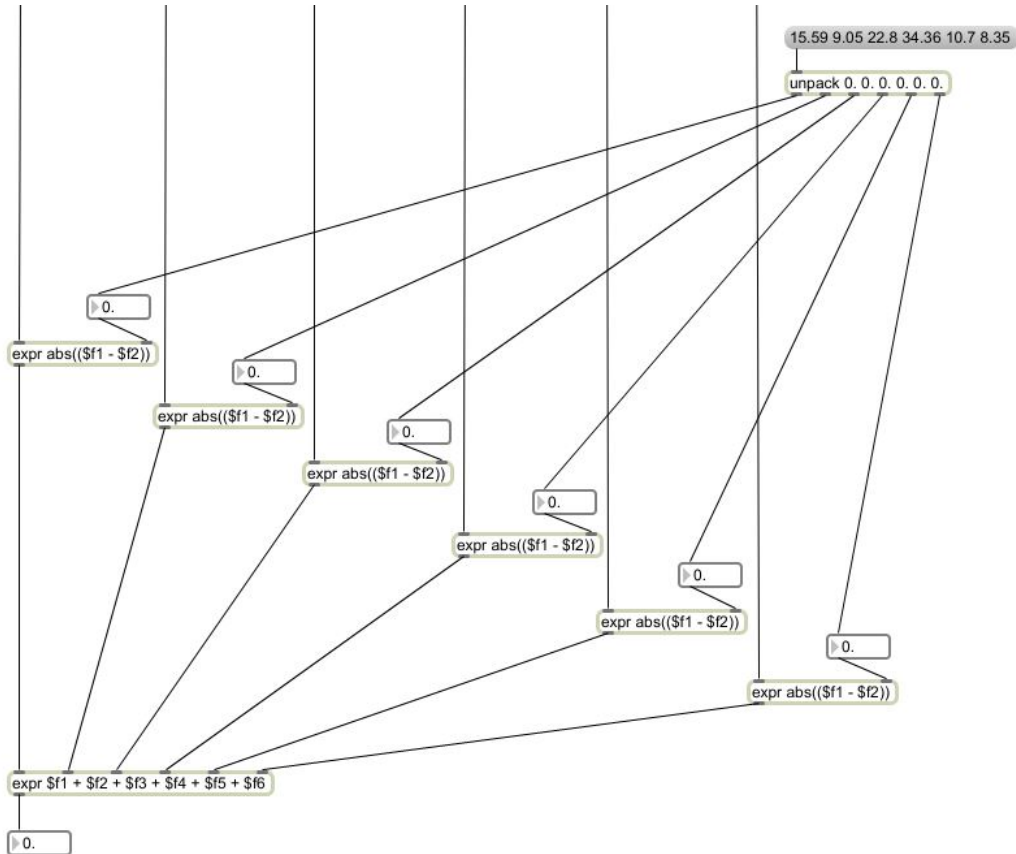


Figura 6: Schema di valutazione dell'errore complessivo.

NOTA: Due componenti *message*, contenenti i valori degli array sperimentali, sono mostrati all'utente. É indispensabile fare click su entrambi per caricare nello schema di valutazione dell'errore i valori statici relativi alle vocali note.

2.7 Decisione per massima verosimiglianza

Una volta valutato il coefficiente di errore rispetto all'una e all'altra vocale, decidere a quale ci si avvicina maggiormente è un'operazione molto semplice.

Un blocco *if* controlla se la sottrazione dell'errore rispetto ad "U" all'errore rispetto ad "A" è un numero positivo o negativo: nel primo caso si decide quindi per la vocale

"A", nel secondo per "U". Un messaggio contenente la vocale corretta viene inviato ad un componente che permette di mostrarlo all'utente.

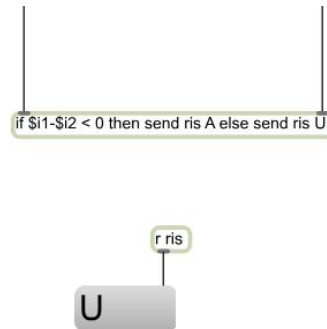


Figura 7: Decisione per massima verosimiglianza.

3 Conclusioni

3.1 Risultati ottenuti

Il risultato ottenuto è, per gli obiettivi prefissati, soddisfacente.

L'input analizzato, che provenga da file o da microfono, produce valori coerenti con la previsione teorica, e l'utilizzo di tali valori per pilotare gli oscillatori del sintetizzatore permette di verificare l'efficacia di quest'ultimo.

Il suono sintetizzato non è fedele a quello analizzato; l'analisi di strumenti musicali, ad esempio, crea set di dati con i quali il sintetizzatore genera suoni molto distanti dal modello. É tuttavia possibile riconoscere facilmente se il suono generato è stato sintetizzato sulla base di dati raccolti dall'analisi di una vocale specifica, e di quale vocale si tratti.

Riguardo alla caratterizzazione della vocale in input, se i dati non sono sufficienti per riprodurre il suono fedelmente, lo sono sicuramente per decidere in maniera precisa se si tratta di "A" o di "U".

3.2 Possibili migliorie e sviluppi

Sarebbe certamente opportuno estendere il riconoscimento dei suoni a tutte le vocali cantate, e tentare di applicare lo stesso principio all'individuazione per timbro di diversi strumenti musicali.

Riguardo alla sintesi, sarebbe interessante ampliare numericamente il set di oscillatori, al fine di riprodurre i suoni ad un livello di fedeltà tale da permettere un processo di analisi e sintesi efficace per scopi di riproduzione musicale.