

POLITECNICO DI TORINO

Collegio di Ingegneria Informatica, del Cinema
e Meccatronica

**Corso di Laurea Magistrale in
Ingegneria Informatica (Computer Engineering)**

Tesi di Laurea Magistrale

Ad Replacement: Progettazione e sviluppo di un sistema di sostituzione delle pubblicità nello streaming digitale di una radio



Relatore

Prof. Giovanni Malnati

Candidato

Alfonso Lentini

Ottobre 2017

Titolo

Ad Replacement: progettazione e sviluppo di un sistema di sostituzione delle pubblicità nello streaming digitale di una radio

Autore

Alfonso Lentini

Relatore

Prof. Giovanni Malnati

Dipartimento di Automatica e Informatica
Politecnico di Torino

Sommario

La rivoluzione digitale ha modificato profondamente il modo con cui le persone comunicano tra loro, abilitando comportamenti ed interazioni su vasta scala fino a poco tempo fa impensabili.

Questo ha avuto profonde ripercussioni anche sul mondo dei media tradizionali, che da strumenti di comunicazione di massa prevalentemente unidirezionali si stanno trasformando in comunità interconnesse di persone che interagiscono tra loro.

Uno dei settori in cui tale impatto risulta evidente è quello delle radio e delle televisioni: sempre più le trasmissioni coinvolgono gli utenti finali che diventano generatori di contenuti che vengono poi diffusi alla collettività tramite il canale radiofonico. Al di là dell'impatto comunicativo, le ripercussioni della rivoluzione digitale si presentano sul piano del modello economico di riferimento. Nel caso specifico delle radio locali, la principale fonte di guadagno è tradizionalmente costituita dalla raccolta pubblicitaria che, per la natura del mezzo, risulta spesso generalista nei contenuti, non potendo rivolgersi a segmenti troppo specifici. La possibilità di distribuire i contenuti radiofonici in streaming apre l'opportunità di personalizzare i contenuti pubblicitari sui singoli utenti, debitamente profilati. Sebbene questo sia diventato il principale approccio per quei media, prevalentemente statici come i giornali, in cui la fruizione avviene in differita, questo approccio non è ancora diffuso nelle trasmissioni in diretta, in cui l'esigenza di prestazioni e la scalabilità dei sistemi pone sostanziali limiti sulla personalizzazione spinta dei contenuti.

Questa tesi ha studiato e prototipato una soluzione scalabile e a costo limitato per la personalizzazione degli annunci in diretta, compatibile con il modello di business e l'infrastruttura tecnologica di quelle radio di piccole e medie dimensioni che, per la loro natura, non possono permettersi grandi investimenti sul piano delle apparecchiature e della rivisitazione dei processi interni. Obiettivo della tesi è stato indagare e mettere a punto una soluzione a basso costo, a livello complessivo di sistema, in grado di fornire flussi audio in diretta integrati con contenuti personalizzati sul profilo del singolo ascoltatore. Tra i requisiti fondamentali che sono stati posti nell'affrontare il problema, ci sono i seguenti:

- l'ascoltatore deve ricevere un flusso continuo, senza pause o interruzioni, corrispondente alla trasmissione in diretta o a inserzioni pubblicitarie personalizzate quando questa si interrompe;
- l'ascoltatore deve poter iniziare la riproduzione in un momento qualsiasi, anche durante la trasmissione di contenuti pubblicitari;
- il meccanismo di riproduzione deve essere sufficientemente semplice e compatibile con le risorse e le architetture dei dispositivi mobili, deve poter certificare l'avvenuta riproduzione delle inserzioni personalizzate e non deve avere impatto significativo sul consumo energetico del dispositivo;
- la durata e la collocazione temporale degli stacchi pubblicitari non è nota in anticipo, ma deve essere rilevata attraverso un opportuno meccanismo;
- il meccanismo di identificazione degli stacchi non deve comportare modifiche sostanziali all'architettura di trasmissione del segnale né deve richiedere un continuo aggiornamento informativo da parte dell'emittente stessa;
- il sistema deve essere sufficientemente generico da poter essere compatibile con una molteplicità di infrastrutture di trasmissione, così da permetterne l'implementazione su una pluralità di stazioni.

A tale scopo sono state indagate ed implementate due diverse soluzioni: la prima basata su marcatori ad ultrasuoni, che vengono sovrapposti alle tracce audio per identificare i segmenti pubblicitari e la loro durata; la seconda basata sull'utilizzo di metadati. Per soddisfare il vincolo di assenza di interruzioni, l'architettura di riproduzione pensata in entrambi i casi è basata sulla presenza di due processi di riproduzione all'interno dei dispositivi degli utenti: uno destinato al flusso in diretta, l'altro alla riproduzione di annunci personalizzati. Il primo processo è sempre attivo e contiene, al proprio interno, la funzione di identificazione dei marcatori. Quando uno di essi viene rilevato provvede ad attivare il secondo processo ed effettuare un cross-fading tra il segnale corrente e l'altro. Quest'ultimo deve poter cominciare la riproduzione con un tempo di intervento inferiore ai 100 ms. Poiché tale valore non è in generale compatibile con le tipiche latenze introdotte dalle diverse reti, il sistema progettato si basa sul pre-caricamento di un certo numero di inserzioni che vengono salvate localmente in un opportuno buffer e di qui prelevate quando servono. Poiché la durata degli stacchi non è nota a priori e non è, in generale, un multiplo esatto della durata delle singole inserzioni, occorre che il sistema di marcatura fornisca un'indicazione della durata complessiva oppure che sia presente un secondo marcatore che identifichi la fine dello stacco. Nel primo caso, il sistema di riproduzione seleziona dal buffer un numero opportuno di inserzioni, tali da coprire per difetto la durata complessiva dello stacco. L'eventuale tempo mancante verrà riempito con un jingle musicale ripetuto quanto serve. Nel secondo caso, verrà finita la riproduzione dell'inserzione corrente e successivamente ripresa in leggera differita la riproduzione del flusso principale; al successivo stacco pubblicitario, questo verrà recuperato ritardando la trasmissione del successivo annuncio di un tempo pari a quello in eccesso dovuto allo stacco precedente. L'approccio basato sull'uso di marcatori ad ultrasuoni prevede l'utilizzo di segnali ad alta frequenza. La scelta di questi segnali nasce dalla teoria dei DTMF (Dual-Tone Multi-Frequency), la quale prevede che un

segnale, per essere facilmente identificabile, possa essere composto dalla somma di due opportune frequenze, una *bassa* e una *alta*. Allo stesso modo sono stati generati segnali composti da una frequenza molto bassa (3hz) e una molto alta ($\sim 18\text{kHz}$) in modo da essere non udibili; questa soluzione tuttavia non è stata adottabile per via dei filtri presenti nelle classiche architetture radio che tagliano le basse frequenze. Per questo motivo si è scelto di utilizzare segnali composti esclusivamente da alte frequenze in sequenza temporale e non sovrapposte perché la sovrapposizione di due alte frequenze può portare alla presenza di subarmoniche udibili all'orecchio umano. Ogni segnale è quindi la composizione di quattro onde sinusoidali ognuna della durata di 250ms. Questi vengono inseriti come marcatori nel secondo precedente l'inizio di uno spot e nel secondo che ne precede la fine, andando a sommarli ai jingle obbligatori per legge inseriti prima e dopo di ogni stacco pubblicitario. In questo modo, il sistema di riproduzione è in grado di reagire con una differenza di 100ms.

Diversamente, l'approccio basato sull'uso di metadati utilizza delle informazioni inserite all'interno della traccia digitale e codificate nel flusso audio da un encoder pilotato per mezzo di una scheda a contatti. Il software di regia, al momento opportuno, genera un segnale che chiude un contatto il quale segnala all'encoder l'iniezione di un prestabilito metadato. In figura 1 viene rappresentata la catena architetturale modificata in modo da permettere questo funzionamento. Allo stesso modo della soluzione precedente, l'errore di marcatura è dell'ordine di 100 ms.

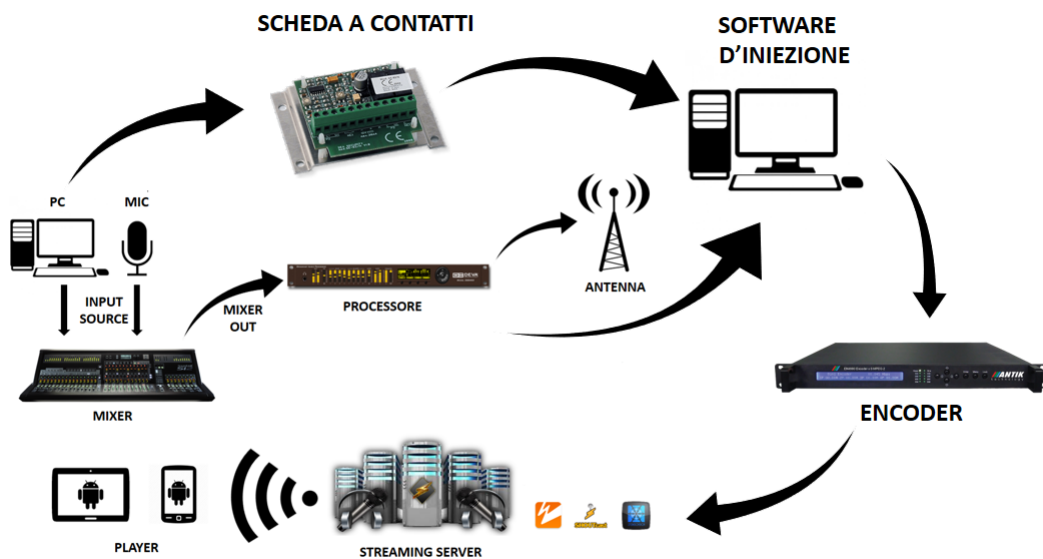


Figura 1: Architettura modificata per l'iniezione di metadati

Il lavoro effettuato va dalla progettazione del sistema d'inserimento dei marcatori fino all'implementazione di due prototipi di applicazioni per dispositivi android, in grado di riprodurre lo streaming digitale e sostituire gli spot trasmessi con altri personalizzati sull'utente. Sulla base di tali implementazioni, sono stati effettuati test con utenti provenienti da emittenti radiofoniche volti a valutare l'accettabilità della soluzione proposta non solo sul piano architetturale ma anche su quello percettivo. Tali test hanno superato le aspettative iniziali certificando l'accettabilità dei tempi d'intervento del sistema e la non udibilità, anche per ascoltatori esperti, del segnale modificato.

*Alla mia famiglia,
solido pilastro della mia vita.*

*"Program testing can be used to show the presence of bugs,
but never to show their absence"*

Edsger W. Dijkstra

Ringraziamenti

Scrivere i ringraziamenti di questa tesi è un passo importante, nel farlo spero di non dimenticare nessuno.

In primis vorrei ringraziare il prof. Giovanni Malnati, non solo per la fiducia accordatami per questo lavoro di tesi ma soprattutto per avermi insegnato, nei corsi tenuti e per insegnarmi quotidianamente questa professione con costanza e fiducia.

Un grande ringraziamento va a Jetmir, Eugenio, Francesca Bobba, Francesca Mondelli, Jessica, Pasquale, Federica, Shoya, Luigi, Andrea pocos e Giusbo per avermi aiutato e motivato durante questo periodo di tesi e infine per avermi accolto in quella che io considero una grande famiglia chiamata TonicMinds.

Un doveroso ringraziamento va ovviamente a tutta la mia famiglia.

A mia madre consegno virtualmente questa mia ultima fatica in segno di riconoscimento per gli sforzi sostenuti, non solo economici. Spero tu sia fiera di aver cresciuto un figlio come me. Ti voglio tanto bene.

A mia nonna e mio nonno, per gli insegnamenti che hanno saputo darmi nel corso degli anni e per quanto di buono mi hanno trasmesso, senza dimenticare tutte le preghiere a Santa Rita, ormai anche la Santa sa programmare !! Vi voglio bene.

A mia sorella, che negli ultimi anni, detto ingegneristicamente parlando, la distanza che ci separa fisicamente è stata inversamente proporzionale a quella nel cuore.

A zio Giuseppe che fin da piccolo mi ha inculcato la passione per l'informatica, grazie a lui ho scoperto questo mondo di cui non mi sono mai pentito.

Il ringraziamento più caloroso è per Valeria che con amore, pazienza e fiducia mi ha sostenuto in questi ultimi anni. Con la speranza che continui ad essere sempre al mio fianco, nei momenti belli e in quelli brutti.

Il ringraziamento storico va al mio cuginazzo Claudio, storico perché lui aveva già previsto, nel lontano 2006, che sarei diventato ingegnere al politecnico di Torino, quando, presentandomi ad alcune sue "colleghe", mi ha definito uno studente di ingegneria al politecnico di Torino. Beh cugino, non potevo mica farti passare per bugiardo quindi lo sono diventato davvero !

Ai miei cugini Francesco e Michele (e anche Claudio) che mi hanno sempre defi-

nito un genio dell'informatica mentre io dentro di me pensavo "ma per quale valido motivo pensano ciò di me? faccio solo cazzate che saprebbe fare chiunque"; beh, sappiate che la spinta nel migliorarmi sempre di più è venuta anche grazie a voi, per non smentire le vostre parole e non deludervi mai.

A zia Rosetta e zio Aldo, che quando ero allo sbando mi hanno accolto quotidianamente nella loro casa e mi hanno rimesso sulla retta via. Come dici tu zia *Ad maiora*, ti mando un bacio zio.

Indice

1	Introduzione	15
1.1	Prefazione	15
1.2	Obiettivi	15
1.3	Organizzazione della tesi	16
2	Definizione del problema	17
2.1	Storia della pubblicità radiofonica	17
2.2	Formati utilizzati	18
2.2.1	Durata degli spot	18
2.3	Trasmissioni streaming	19
2.4	Il mercato dello streaming audio	20
2.4.1	Audience	21
2.5	Servizi internet radio	22
2.5.1	Come le persone ascoltano le internet radio	22
2.6	Gestione delle inserzioni	24
2.7	Sistema di messa in onda	24
2.8	Stato dell'arte	26
2.8.1	Profilazione degli utenti	27
2.8.2	Tecnologie d'inserimento	30
2.9	Motivazioni	32
3	Processamento del segnale audio	33
3.1	Musica e fisica	33
3.1.1	Toni puri e suoni reali	33
3.2	Digitalizzazione	35
3.2.1	Campionamento	35
3.2.2	Quantizzazione	37
3.2.3	Pulse Coded Modulation	38
3.3	Dal suono digitale alle frequenze	38
3.3.1	Trasformata discreta di Fourier	39
3.3.2	Fast Fourier Transform	40
3.4	Algoritmo di Goertzel	41
3.4.1	Funzionamento dell'algoritmo base	41
4	Metodologia	44
4.1	Sistemi di comunicazione	44
4.1.1	Sistemi di comunicazione digitale	46
4.2	DTMF	47
4.3	Audio Metadata	48

4.4	Audio encoder	49
4.4.1	Descrizione dell'algoritmo	51
4.4.2	Strumenti per migliorare l'efficienza di codifica	53
5	Sviluppo del lavoro	56
5.1	Implementazione tramite metadati	56
5.1.1	Architettura del sistema	56
5.1.2	Funzionamento	59
5.1.3	Applicazione android	61
5.2	Implementazione tramite ultrasuoni	65
5.2.1	Architettura del sistema	65
5.2.2	Funzionamento	66
5.2.3	Applicazione android	70
6	Conclusioni	74
6.1	Risultati, vantaggi e svantaggi	74
6.2	Sviluppi futuri	75
6.2.1	Podcast	75
6.2.2	Sincronizzazione in-door	76

Elenco delle figure

1	Architettura modificata per l'iniezione di metadata	4
2.1	Report settimanale degli ascoltatori di internet radio, condotto dal Edison Research nel 2014	21
2.2	Report di Edison Research su quali tipi di audio ascoltano gli americani	21
2.3	Report di Triton Digital's Webcast Metrics sullo streaming audio su dispositivi mobili	23
2.4	Report di comScore e Millennial Media sul tempo trascorso sulle internet radio in diversi dispositivi	23
2.5	Modello di un sistema di messa in onda di una radio	25
2.6	Schema a blocchi di un processore audio	25
2.7	Principio base del metodo collaborativo	29
3.1	Rappresentazione di un'onda sinusoidale pura di frequenza 20hz e di ampiezza 1	34
3.2	Rappresentazione di una composizione di onde sinusoidali	34
3.3	Rappresentazione di un onda sinusoidale a 20 hz campionata a 40hz .	36
3.4	Rappresentazione di un onda sinusoidale a 20 hz campionata a 30hz .	36
3.5	Rappresentazione di un segnale quantizzato a 8 livelli	37
3.6	Rappresentazione di un segnale quantizzato a 64 livelli	38
3.7	Rappresentazione binaria di un campione PCM 16 bit stereo	39
4.1	Blocchi fondamentali costituenti un sistema di comunicazione	45
4.2	Blocchi fondamentali di un sistema di comunicazione digitale	46
4.3	Keyboard layout DTMF	47
4.4	Diagramma a blocchi di un sistema di codifica/decodifica percettiva. .	50
4.5	Diagramma a blocchi di un MPEG-1 Layer-3 encoder.	51
4.6	Diagramma a blocchi di un MPEG-2 AAC encoder.	53
5.1	Software di regia Mb Studio	57
5.2	Architettura modificata per l'iniezione di metadata	58
5.3	Finestra d'impostazione bit MbStudio	58
5.4	Esempio di palinsesto nel software MbStudio	59
5.5	Esempio d'impostazione bit nel software MbStudio	60
5.6	User interface per la selezione del profilo	61
5.7	Flow chart download spot pubblicitari	62
5.8	User interface di riproduzione Exoplayer	63
5.9	Stati del player (a) programma live (b) programma differita (c) spot personalizzato	64
5.10	Grafico dei volumi durante il crossfade	65

5.11	Architettura sistema di messa in onda	66
5.12	Forma d'onda di un segnale a 18 khz	68
5.13	Palinsesto con jingle/segnale d'inizio spot	68
5.14	Forma d'onda del segnale composto con gli ultrasuoni	69
5.15	Diagramma a blocchi di un MPEG-1 Layer-3 encoder.	69
5.16	User interface per il login social.	70
5.17	Rappresentazione del flusso di decodifica.	71
5.18	Macchina a stati del SequenceChecker.	72
5.19	User interface del player principale.	72
6.1	NodeMcu con amplificatore e speaker ad ultrasuoni	76

Elenco delle tabelle

2.1	Confronto tra i tipi di profilo	28
2.3	Confronto tra i vari metodi	30
5.1	Frequenze scelte in Hz	67
5.2	Segnali generati a partire dalle frequenze scelte	67

Capitolo 1

Introduzione

1.1 Prefazione

Il presente lavoro ha come oggetto lo sviluppo di tecniche per l'Ad Replacement (sostituzione delle pubblicità) nella radio digitale.

Verranno presentati due differenti approcci al problema; il primo si concentrerà sull'utilizzo dei metadati, l'altro sull'utilizzo degli ultrasuoni.

Entrambi dovranno fungere da canali di comunicazione in modo da trasmettere l'informazione sull'arrivo e la fine di una pubblicità. Nel primo caso la comunicazione avviene tramite l'iniezione di tag metadata nel flusso audio, nel secondo tramite l'iniezione di ultrasuoni. I metadati, anche comunemente denominati metadati ID3, sono informazioni incorporate in un file audio utilizzate per identificarne il contenuto. Gli ultrasuoni sono vibrazioni meccaniche della materia che si trasmettono sotto forma di segnali sonori, la cui principale caratteristica consiste nel fatto che l'udito umano non le può captare perché hanno una frequenza superiore a 18 Khz.

Gli ultrasuoni trovano utilizzo in molteplici applicazioni, possono misurare le distanze, localizzare gli ostacoli o permettere la comunicazione fra un robot e il suo controllo o fra vari robot. La maggior parte di queste applicazioni, sfrutta una delle caratteristiche fondamentali degli ultrasuoni: la riflessione.

Ciò nonostante, all'interno di questo lavoro, verrà sfruttata un'altra proprietà degli ultrasuoni, la frequenza e la sua intensità misurata in decibel. Come detto precedentemente, gli ultrasuoni sono onde la cui frequenza è superiore ai 18-20 Khz. Proprio grazie alla varietà di frequenze che si possono incontrare in questo intorno si possono generare segnali diversi da utilizzare come codice¹ di un messaggio inviato da un *emittente* a un *ricevente*.

1.2 Obiettivi

Lo scopo finale di questa tesi è lo sviluppo di applicazioni per dispositivi Android, in grado di captare segnali, siano essi metadati o ultrasuoni, e reagire al fine di implementare un sistema *"on the fly"*.

Con il termine *"on the fly"* si vuole descrivere qualcosa che sta per cambiare mentre il processo che subisce questo cambiamento è in corso.

¹Codice: parola pronunciata o scritta, immagine, tono impiegato per "formare" il messaggio della comunicazione.

Nel caso in studio, uno dei processi che si è voluto far mutare è la riproduzione dello streaming audio di un'applicazione per la riproduzione di una web radio, in conseguenza alla trasmissione di una pubblicità.

Nello specifico, l'applicazione deve essere in grado di analizzare il flusso di streaming ed individuare l'arrivo di una pubblicità, o tramite i segnali ultrasonici o tramite metadati, iniettati dall'emettitore nel flusso audio, sostituendola con un'altra mirata sugli interessi dell'ascoltatore.

Un altro campo nel quale si è applicato questo sistema di segnalazione è quello dei musei; tramite l'utilizzo di microcontrollori posti su ogni opera facenti funzione di emettitori ultrasonici, è stato possibile creare un'applicazione che tramite il microfono, all'avvicinarsi dell'utente all'opera, captasse questi segnali e interagisse in automatico con l'utente fornendo informazioni riguardo all'opera che si trova di fronte.

1.3 Organizzazione della tesi

Nel capitolo 2 viene descritto il problema alla base del lavoro di questa tesi che ha portato ad utilizzare la tecnologia ad ultrasuoni come possibile soluzione; inoltre, vengono descritte le tecnologie utilizzate al momento.

Nel Capitolo 3 viene esposta la teoria alla base del processamento audio, dalla digitalizzazione ai teoremi e le operazioni matematiche utilizzate per l'analisi dei segnali.

Nel capitolo 4 vengono presentate alcune tecnologie utilizzate nel successivo capitolo per l'implementazione delle soluzioni trovate.

Nel capitolo 5 viene presentato lo sviluppo del lavoro effettuato, illustrando passo dopo passo le tecniche utilizzate mettendole a confronto.

Infine, nel capitolo 6, vengono presentate le conclusioni e i possibili sviluppi futuri delle tecniche utilizzate.

Capitolo 2

Definizione del problema

In questo capitolo viene esposto da dove nasce la necessità di implementare queste tecniche di comunicazione; il principale problema che ha portato allo sviluppo di questo lavoro, è stato quello di come fare ad inserire, nel flusso streaming di una web radio, inserzioni pubblicitarie personalizzate sull'ascoltatore.

Nei prossimi paragrafi viene esposto il contesto delle web radio da cui è possibile trarre le motivazioni alla base di questa ricerca, infine verranno brevemente descritte le principali tecnologie utilizzate al momento.

2.1 Storia della pubblicità radiofonica

Le prime trasmissioni radio furono mandate in onda nei primi anni del 1900, tuttavia, fino al 1919 le stazioni radio non diffondevano messaggi pubblicitari.

Nel 1919 l'ingegnere Frank Conrad fu il primo a pubblicare un annuncio radiofonico sulla propria stazione sperimentale nel quale ringraziava un negozio di musica di Pittsburgh per avergli fornito registrazioni fonografiche.

Più in là i titolari di stazioni radio si trovarono sempre più a fronteggiare la questione di come sostenere finanziariamente le loro stazioni, perché operare una stazione radio era una spesa significativa. Nel febbraio del 1922, AT&T annunciò che avrebbe iniziato a vendere "toll broadcasting" agli inserzionisti le cui imprese avrebbero sottoscritto o finanziato una trasmissione, in cambio di essere citati in radio.

Durante l'età d'oro della radio, gli inserzionisti hanno sponsorizzato interi programmi, di solito con qualche tipo di messaggio, come "ringraziamo i nostri sponsor per rendere possibile questo programma", in onda all'inizio o alla fine di un programma. La pubblicità era diventata una merce calda e c'erano soldi da fare.

L'industria radio-televisiva è cambiata significativamente da quella prima trasmissione nel 1919 e la radio oggi è un grande business. Sebbene oggi altri media e nuove tecnologie sono a disposizione del consumatore, il 95% delle persone ancora ascolta la radio ogni settimana[9]. Anche l'ascolto radiofonico in Internet sta crescendo, inoltre, sebbene i consumatori abbiano più scelte, oggi il 92% rimane in sintonia quando le pubblicità entrano nella loro programmazione.

2.2 Formati utilizzati

Con i cambiamenti nell'industria radiofonica e le migliori tecnologie di produzione, il modo di presentazione commerciale è cambiato e gli annunci possono assumere un'ampia gamma di forme. I due tipi principali di annunci radio sono i "live reads" e i "produced spots".

I "live reads" sono quegli annunci che il dj legge da un foglio informativo durante una diretta radio. I "produced spots", che sembrano essere più comuni, sono annunci registrati dalla stazione radio o da un'agenzia pubblicitaria per conto di un cliente. Essi includono lettura diretta con effetti sonori o musica di sottofondo, dialoghi, monologhi, jingle e combinazioni di questi. Gli studi dimostrano che la qualità degli spot è altrettanto importante per gli ascoltatori, così come il numero di annunci che ascoltano.

Le stazioni radio oggi gestiscono generalmente la loro pubblicità in cluster o in set, sparsi per tutta la durata della trasmissione. Gli studi dimostrano che il primo o il secondo annuncio commerciale durante un'interruzione ha un richiamo più elevato rispetto a quelli che si trasmettono più tardi nell'insieme[2].

Conoscere le classifiche delle radio più ascoltate, il target demografico e avere valutazioni su esse diventa quindi fondamentale per l'inserzionista. A questo scopo, le agenzie pubblicitarie si iscrivono a servizi a pagamento, come Nielsen Audio [25] negli USA, per ottenere queste classifiche in modo da selezionare un segmento specifico del pubblico di ascolto, ad esempio, un loro cliente potrebbe voler raggiungere gli uomini dai 18 ai 49 anni.

I numeri possono mostrare chi sta ascoltando una stazione particolare, i momenti più popolari del giorno per gli ascoltatori di quel gruppo e la percentuale del pubblico di ascolto totale raggiungibile con un particolare programma di pubblicità. I numeri mostrano esattamente quante persone ascoltano in ogni ora del giorno. Ciò consente ad un inserzionista di selezionare le stazioni più forti sul mercato e scegliere quali orari del giorno saranno i tempi migliori per eseguire i propri annunci.

Le stazioni radio vendono il loro tempo di trasmissione secondo fasce orarie giornaliere. Le fasce in cui le persone sono in viaggio, sono solitamente i tempi più apprezzati della giornata poiché sono i momenti in cui ogni stazione ha più ascoltatori. Le tariffe o quello che la stazione radio addebita all'inserzionista, ne rifletterà. Pertanto, i tassi di pubblicità variano a seconda del giorno dell'anno, dell'ora del giorno, della stabilità della stazione o dalla posizione demografica degli ascoltatori. Inoltre i tassi di pubblicità variano a seconda della durata che gli inserzionisti scelgono di eseguire. Sebbene siano i sessanta secondi di spot i più comuni, le stazioni vendono anche tempi di trasmissione di trenta, quindici, dieci e due secondi.

2.2.1 Durata degli spot

All'inizio del 2005, Clear Channel Communications ha portato l'industria radiofonica in un nuovo mondo[35]. Per la prima volta, gli annunci di 30 e 15 secondi avevano un prezzo che valeva la pena. Fino ad allora, tutti gli annunci radio avevano essenzialmente lo stesso prezzo, indipendentemente dalla lunghezza, per cui tutti utilizzavano il formato da sessanta secondi e gli ascoltatori erano costretti a trascorrere troppo tempo ad ascoltare gli annunci.

Con l'iniziativa "Less is More", Clear Channel negli Stati Uniti, ha iniziato ad usare trenta secondi di spot. Sebbene gli studi dimostrano che meno pubblicità comporta una maggiore frequenza di richiamo[34], la ricerca suggerisce che tradizionali spot di sessanta secondi potrebbero essere l'opzione migliore, con una maggiore ricorrenza di marca e di messaggi rispetto ai nuovi annunci di trenta secondi. Inoltre le stazioni eseguono anche spot di dieci secondi; tipicamente questo tipo di spot si esegue adiacente ad alcune funzionalità delle stazioni, come il rapporto sul traffico, indicando "questo viene fornito per voi da ..." e di solito è limitato a circa trenta parole. Spot di quindici secondi infine sono generalmente riservati agli annunci promozionali della radio, anche se alcune stazioni li vendono[35].

2.3 Trasmissioni streaming

Oltre alla pubblicità radio tradizionale, alcune stazioni stanno vendendo spazi pubblicitari durante le trasmissioni in streaming. In passato, le stazioni radio includevano solo le pubblicità che erano in esecuzione anche "On Air" fino a che la Columbia Broadcasting System annunciò che avrebbe iniziato a trasmettere "live reads" nelle sue trasmissioni radiofoniche in streaming, vendute separatamente dai regolari spot delle stazioni, notandone l'efficacia.

La radio Internet ha il potenziale di essere la forma più onnipresente dei media in assoluto, persino più dei film, della televisione o dei libri. Questo perché l'ascolto di musica può essere goduto mentre si svolgono altre attività. Prendendo in considerazione il potenziale delle internet radio, si prevede che in due anni 3,5 miliardi di persone saranno in linea, portando le possibili ore di ascolto a livello mondiale a 59,5 miliardi al giorno. Con il reddito medio per mille ore pari a \$ 42,77 (tasso di Pandora¹ nel 2014), già nel 2017 è possibile un capitale giornaliero di circa 2,5 miliardi di dollari. Questo presuppone che il tasso di mercato sia uguale in tutto il mondo, che al momento non è il caso.[29]

Per poter realizzare questo potenziale, le aziende devono fornire esperienze di ascolto altamente personalizzate che devono ancora essere pienamente ottimizzate. Questo tipo di personalizzazione è già iniziata nella pubblicità, ma non è stata ben implementata nelle internet radio. Tuttavia, le barriere della radio in streaming stanno venendo lentamente eliminate. La gamma di dati personali che compongono i gusti delle persone sta crescendo, il che a sua volta significa che possiamo anche capire meglio i gusti di persone simili. Con i costi decrescenti dello streaming, la raccolta e l'archiviazione di grandi quantità di dati, nonché la crescita di potenti strumenti per analizzarla, la nostra capacità di esplorare e trarre inferenze da questa ricchezza di informazioni è apparentemente infinita. Infine, le grandi etichette discografiche sono ora disposte a fare questo spostamento digitale. Negli ultimi anni, le etichette hanno firmato accordi che rendono le loro librerie musicali legalmente accessibili su richiesta, ponendo la base di questa trasformazione nella scoperta digitale basata su grandi dati.

¹Pandora: applicazione per iOS ed Android che permette di ascoltare musica in streaming.

2.4 Il mercato dello streaming audio

L'audio è trasmesso in streaming online ormai da 20 anni. La consapevolezza dei consumatori e l'adozione di audio in streaming sono cresciuti in questo periodo, così come sono cresciute le occasioni per gli inserzionisti.

Le piattaforme audio in streaming hanno generalmente due modelli di business: un modello basato su sottoscrizioni che carica sugli ascoltatori una tassa di accesso oppure un modello basato su annunci che offrono gratuitamente la programmazione agli ascoltatori, monetizzando con le vendite pubblicitarie.

Centinaia di servizi musicali e migliaia di stazioni radio Internet sono oggi sul mercato. La maggior parte degli ascoltatori sceglie l'ascolto gratuito supportato da annunci, rendendo la pubblicità il modello di reddito primario per i publisher. Il pubblico in crescita ha attirato un maggior interesse da parte degli inserzionisti; alcuni lo considerano come un'estensione della pubblicità radio, altri vedono l'audio in streaming come complemento alle campagne pubblicitarie digitali. Sebbene l'audio in streaming prenda in prestito concetti e terminologie chiave dalla radio broadcast (ad esempio "internet radio"), non avviene una traduzione diretta dal lato della programmazione e della pubblicità. L'esperienza del programma è spesso più personalizzata per l'utente, che non di quella "on-air". Dal lato degli affari ci sono differenziatori chiave per gli inserzionisti:

- **Minori inserzioni:** i flussi audio generano meno spot all'ora rispetto alla radio tradizionale, per cui i messaggi di marketing presentano meno concorrenza per l'attenzione dell'ascoltatore. Gli annunci in streaming si distinguono di più.
- **Tecnologia degli annunci digitali:** big data, segmentazione degli spettatori, precisione di monitoraggio, nuovi prodotti pubblicitari - audio in streaming grazie a infrastrutture digitali native. Tutti questi fattori rendono le campagne pubblicitarie streaming più intelligenti ed esatte.

L'audio in streaming è ancora nei primi giorni dal punto di vista dell'opportunità pubblicitaria. Gode della stessa attrazione dei consumatori come la radio broadcast tradizionale: l'amore profondo del mezzo e della programmazione. Allo stesso tempo, lo streaming offre vantaggi convincenti che guidano l'adozione dei consumatori in avanti, fornendo valore e ritorno agli inserzionisti. Nel maggio del 2014, SNL Kagan ha pubblicato l'Economics of Internet Music & Radio, una proiezione di entrate pubblicitarie in streaming dal 2013 al 2023, marcando il livello di Ad-reddito nel 2013 a 889 milioni di dollari, inoltre, SNL Kagan ha previsto una crescita a due cifre, portando il fatturato totale degli annunci in streaming a 1,5 miliardi di dollari nel 2015. Guardando ulteriormente, SNL Kagan ha proiettato il totale delle entrate pubblicitarie (senza contare i webcast terrestri) a 3,2 miliardi di dollari nel 2023. In questo livello previsto, le entrate pubblicitarie dello streaming audio rappresenterebbero il 3,6% del totale degli annunci pubblicitari U.S.[26].

2.4.1 Audience

La forte curva di crescita dei servizi radio e musicali online, è emblematizzata da Pandora Media, il cui servizio di radiodiffusione Internet ha visto un aumento del 112% delle dimensioni del pubblico e un aumento del 283% delle sessioni di streaming dal luglio 2011 al luglio 2014.

Inoltre, la crescita del pubblico per lo streaming audio è diffusa tra molti marchi e piattaforme, promossa dalla tecnologia dei consumatori che alimenta nuovi metodi di ascolto. Non è solo la dimensione del pubblico che sta crescendo; la quantità di tempo trascorso ad ascoltare l'audio in streaming è incrementato di anno in anno.

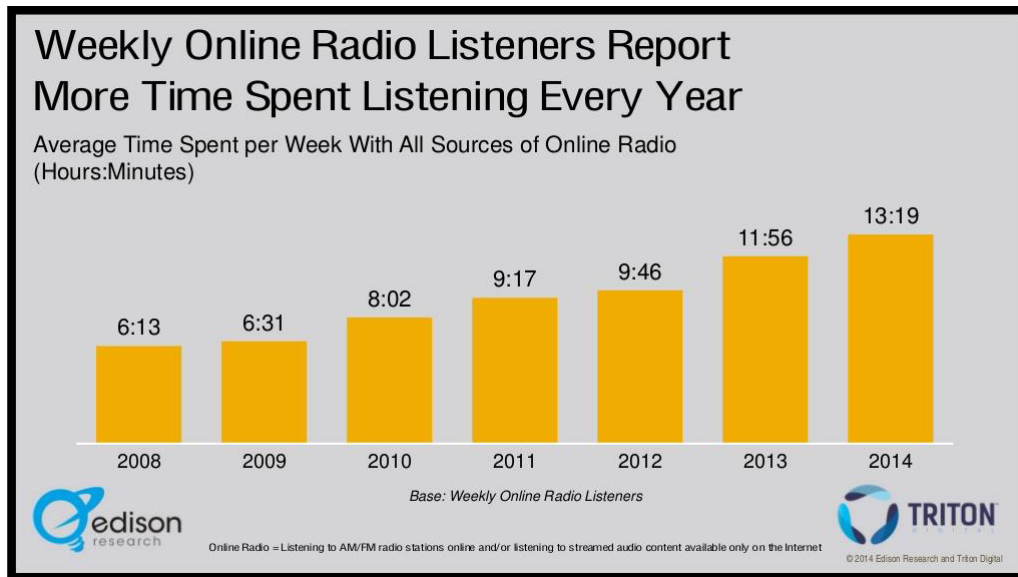


Figura 2.1: Report settimanale degli ascoltatori di internet radio, condotto dal Edison Research nel 2014

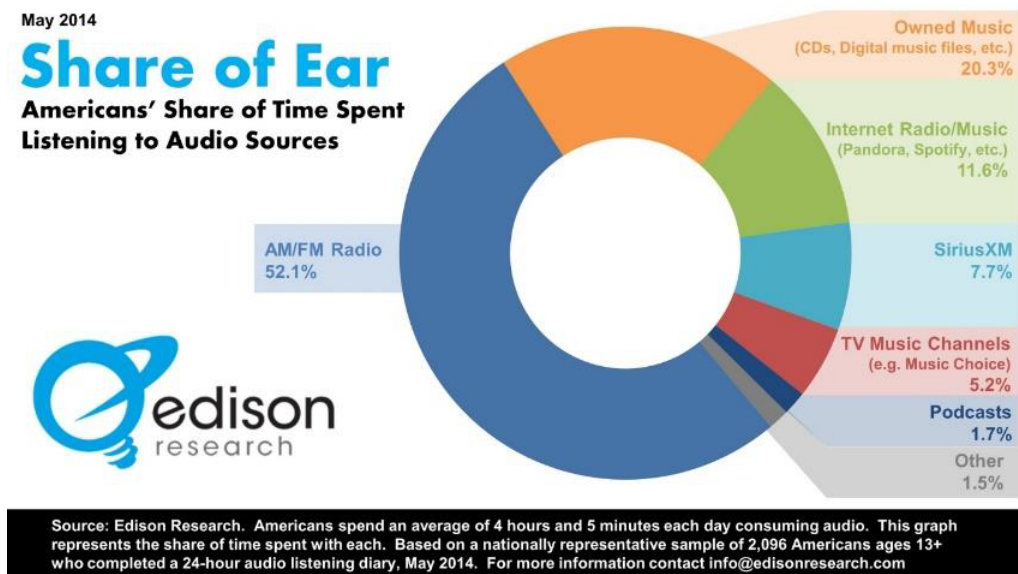


Figura 2.2: Report di Edison Research su quali tipi di audio ascoltano gli americani

2.5 Servizi internet radio

I servizi radio Internet emulano la radiodiffusione trasmettendo un'esperienza di ascolto più limitata. Per definizione, la licenza obbligatoria con la quale operano per accedere alla musica, limita gli elementi interattivi del loro servizio. La maggior parte delle stazioni radio e delle piattaforme Internet dispone di modelli di entrata supportate da annunci. In alcuni casi, viene offerto un piano di sottoscrizione agli utenti, la cui principale caratteristica sta nell'eliminare gli annunci (senza aggiungere funzioni interattive)

Alcune radio Internet sono parzialmente personalizzate. Le internet radio più sofisticate hanno livelli di intelligenza musicale che permettono di conoscere l'utente nel tempo, migliorando l'esperienza di ciascun ascoltatore. La radio di Pandora è il principale esempio di mercato. Alcune internet radio supportate da annunci pubblicitari non contengono nessuna di questa interattività limitata. Migliaia di piccole stazioni online sono più in linea con l'esperienza di trasmissione tradizionale.

2.5.1 Come le persone ascoltano le internet radio

Le persone ascoltano l'audio in streaming su computer, telefoni, tablet, sistemi di altoparlanti wireless, dashboard per auto e attraverso categorie emergenti di dispositivi informatici come i sistemi oculari (ad esempio Google Glass) e altri indossabili come gli smartwatch.

In questo spettacolare paesaggio di dispositivi, l'audio viene trasmesso in streaming direttamente da servizi musicali di marca e applicazioni radio Internet, oppure indirettamente nell'aggregazione di piattaforme che servono più stazioni di streaming. Le persone si connettono ai servizi musicali direttamente dai siti web o tramite app mobili. In un mondo sempre più app-centrico, i siti web dei servizi musicali sono anche chiamati applicazioni web o applicazioni desktop.

Le opportunità di pubblicità e le possibilità creative possono variare da desktop a mobile, a causa delle dimensioni dello schermo drammaticamente diverse. Detto questo, gli annunci audio funzionano ugualmente in tutti i formati dello schermo e poiché l'ascolto dell'audio è in genere un'esperienza off-the-eye, sia in un computer, in una macchina, o in strada con uno smartphone, la pubblicità audio è agnostica. Come per i servizi musicali su Internet, i flussi web della radio terrestre sono (spesso, ma non sempre) disponibili per gli ascoltatori su computer, tablet e telefoni. È comune per i siti web della stazione radio di offrire un pulsante "Ascolta ora" per gli ascoltatori per sintonizzarsi in linea. Nell'implementazione più semplice, il flusso radio è un simulcast esatto del segnale di trasmissione, inclusi gli annunci.

Le stazioni radio terrestri talvolta dispongono di singole applicazioni per dispositivi mobili che forniscono al pubblico il loro streaming. Esistono, infine, piattaforme online che aggregano le stazioni radio; esse servono come directory delle stazioni disponibili e consentono alle stazioni di partecipare facilmente allo spazio di streaming mobile. Il risultato è che le stazioni più terrestri trasmettono anche l'audio in streaming e migliaia di tali flussi vengono consegnati a dispositivi mobili e macchine.

L'audio in streaming è principalmente su piattaforma mobile. Infatti, la internet radio è una delle due attività più diffuse su smartphone, eguagliata solo dai giochi. Nell'edizione di luglio di Triton Digital Webcast Metrics Top 20 Ranker, società di

misurazione per la tecnologia streaming, hanno rilevato che il 68% dello streaming si è verificato sui due principali sistemi operativi mobili - iOS e Android.

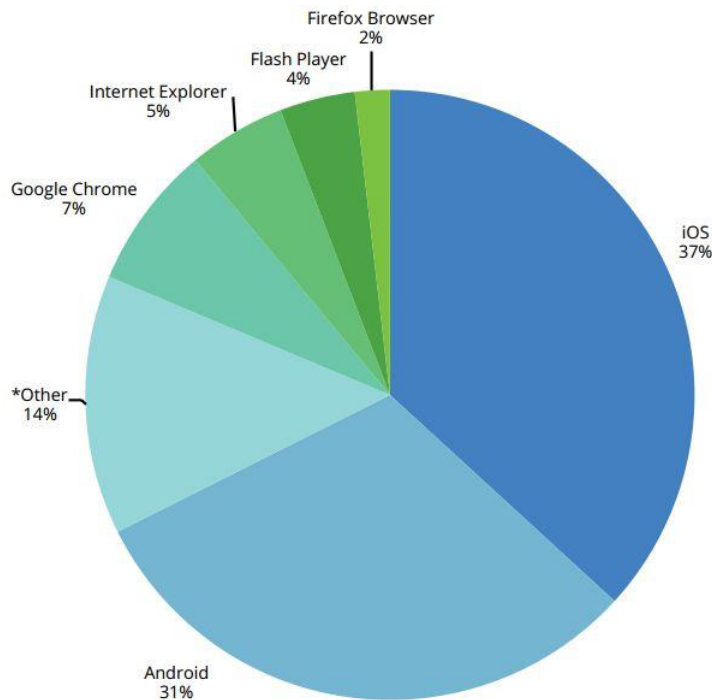


Figura 2.3: Report di Triton Digital's Webcast Metrics sullo streaming audio su dispositivi mobili

Una misura ancora più sorprendente viene da comScore e Millennial Media, che mostra che il 95% del tempo trascorso sulle internet radio nel gennaio 2014 è stato fatto tramite tablet (16%) e smartphone (79%).[24] Il numero combinato è stato superiore a quello di qualsiasi altro tipo di attività online.

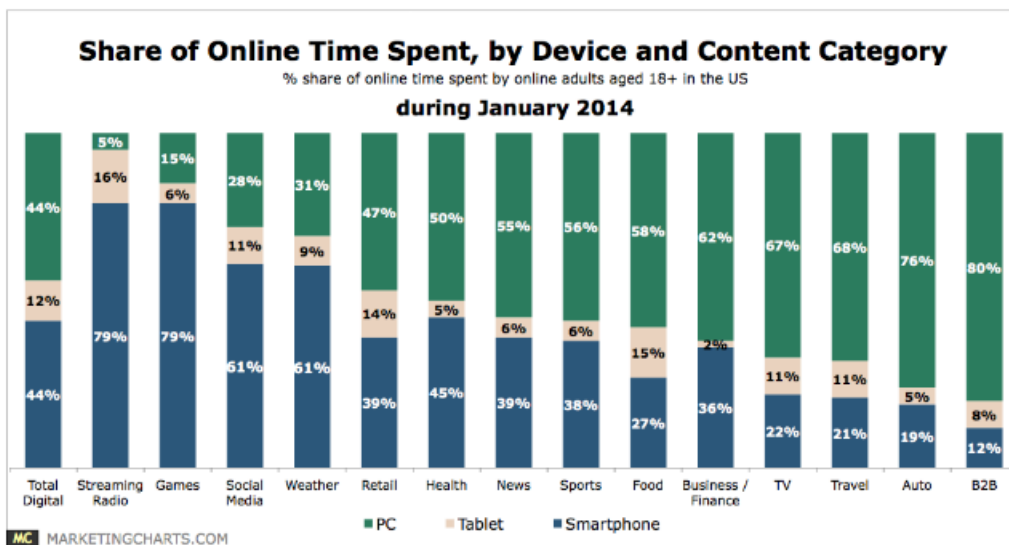


Figura 2.4: Report di comScore e Millennial Media sul tempo trascorso sulle internet radio in diversi dispositivi

Il primato dei dispositivi mobile non è limitato agli Stati Uniti, per esempio quasi il doppio degli intervistati in uno studio in Francia ha ascoltato internet radio attraverso il cellulare piuttosto che attraverso un computer.[14]

2.6 Gestione delle inserzioni

Nel settore dei mass media come quello della radio, il palinsesto è l'insieme delle trasmissioni programmate per un certo periodo (un giorno, una settimana, un mese, un trimestre). Solitamente il palinsesto indica l'ora di messa in onda, il titolo e il tipo di ogni singolo programma, più eventuali informazioni accessorie. Esso è anche definibile come la sequenza temporale dei messaggi offerta dall'emittente a tutti i possessori dell'apparecchio sintonizzati su una data frequenza.

Al suo interno, vengono definiti anche i momenti della giornata in cui andranno in onda gli spot pubblicitari. Questi, sebbene precedentemente programmati, introducono complessità nella gestione del palinsesto in quanto possono variare anche pochi minuti prima della messa in onda. Il motivo di queste variazioni "last minute" possono essere vari, per esempio uno spot può essere programmato alla fine di un programma live dove può capitare che lo speaker termini in ritardo, rubando secondi allo spot, oppure termini in anticipo lasciando secondi di trasmissione vuoti.

Un'altra possibile causa è dovuta ai "remnant advertising", cioè spazi pubblicitari che non sono stati venduti e che possono essere acquistati con uno sconto rapido dell'ultimo minuto. Il tempo pubblicitario è una merce deperibile, se non viene venduto, viene perso. Tuttavia, invece di avere una perdita, le aziende che si occupano della vendita, spesso richiedono molto meno delle loro solite tasse al dettaglio. Ciò significa che gli inserzionisti possono acquistare ciò che tipicamente è costoso per un prezzo molto inferiore di quello normale.

Questa possibile variazione del palinsesto, crea un problema nel momento in cui si vuole segnalare il preciso istante in cui uno spot va in onda, infatti esso non è prevedibile anticipatamente visto che può cambiare all'ultimo minuto.

2.7 Sistema di messa in onda

In figura 2.5 è rappresentato un comunissimo modello di messa in onda di una radio; la figura mostra i vari dispositivi che compongono la catena.

Il segnale prende forma all'interno di una macchina che tipicamente ospita un software specializzato per la regia in cui è definito il palinsesto; il flusso audio generato dal software di regia è combinato all'interno di un mixer con altre sorgenti come, per esempio, il microfono dello speaker.

L'uscita del mixer è generalmente collegata ad un dispositivo chiamato processore, il quale può essere implementato sia in hardware che in software; questo dispositivo è diviso in più blocchi, che svolgono varie funzioni.

La figura 2.6 mostra un esempio di diagramma a blocchi di un processore. In ingresso troviamo un filtro passa alto che consente di rimuovere l'energia subsonica dell'alimentazione prima di procedere con ulteriori elaborazioni. In questo modo, le prestazioni di distorsione dell'intermodulazione del processore e l'audiopath che ne segue sono migliorate e le risorse di elaborazione non vengono sprecate elaborando processi inudibili che non dispongono di informazioni acusticamente rilevanti.

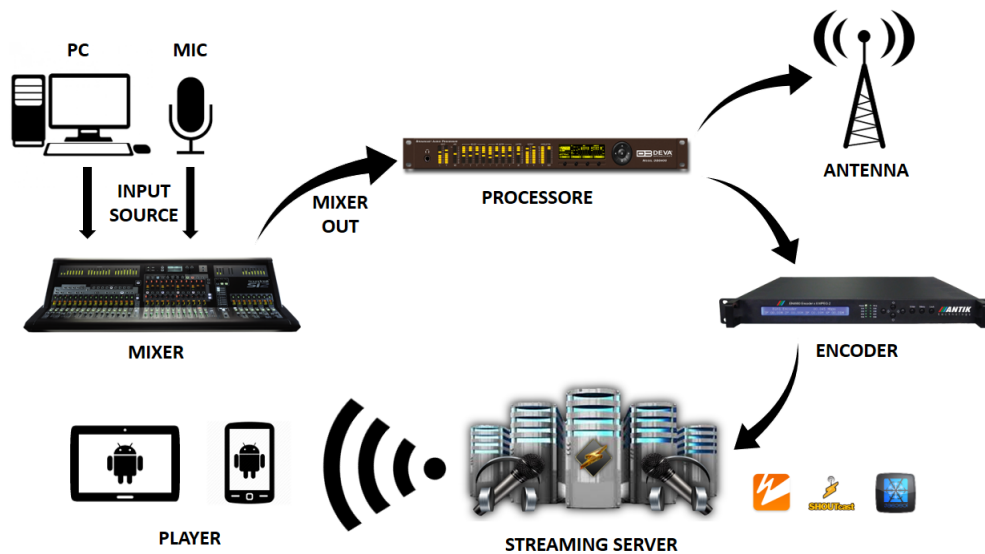


Figura 2.5: Modello di un sistema di messa in onda di una radio

Segue il Wideband AGC (Automatic Gain Control) che svolge un ruolo molto importante. Si tratta di un sistema a circuito chiuso che regola automaticamente il guadagno di tensione in modo che la tensione di uscita rimanga entro un intervallo desiderato. In uscita ad esso viene posto un enhancer bass che esalta le frequenze molto basse

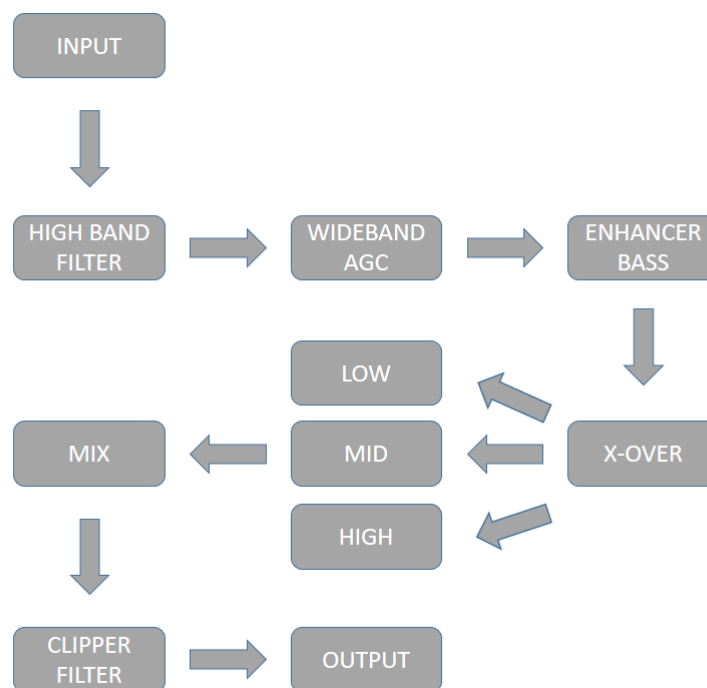


Figura 2.6: Schema a blocchi di un processore audio

come quella del tamburo o note molto basse che sono presenti nel segnale; successivamente si trova l'X-over (Crossover) che suddivide il segnale audio in tre bande

indipendenti per consentire l'elaborazione ottimale per i campi di frequenza bassi (L), medi (M) e alti (H).

Alla fine di questo passaggio il segnale precedentemente splittato dal crossover, viene ricomposto tramite un mixer. Questo è il punto finale prima che il segnale raggiunga il clipper filter il quale controlla la profondità del ritaglio con la precisione di 0.1 dB su una gamma di +/- 6 dB.

A questo punto il segnale in uscita dal processore prende due strade, una porta all'antenna che si occuperà di trasmettere il segnale in onde radio, l'altra porta ad un altro dispositivo, l'encoder.

Uno dei processi fondamentali che si instaurano durante la trasmissione radiofonica sul web è l'encoding. La codifica è quel passaggio che consente di trasformare un segnale di un tipo in un altro che presenta determinate caratteristiche. Negli studi radiofonici assistiamo alla convivenza temporale di segnali aventi caratteristiche differenti, quali la musica, che solitamente è già in formato digitale, e la voce dello speaker. L'encoder non fa altro che convertire e campionare questi segnali eterogenei in un segnale ben definito. Il segnale risultante può essere di diverse tipologie, a seconda che la campionatura venga effettuata da un tipo di algoritmo piuttosto che da un altro o se alla campionatura venga o meno associata una compressione. Esistono quindi una infinità di codifiche audio, esse si dividono in tre macrocategorie: non compressi, compressi con perdite (lossy) e compressi senza perdite (lossless). Tra i compressi con perdite si trovano i codec MP3 e AAC che sono quelli maggiormente utilizzati per lo streaming web. Queste codifiche lossy tendono ad eliminare le frequenze dei suoni impercettibili all'orecchio umano, concentrandosi quindi solo sulla banda di frequenze udibili (50-18000 Hz). Oltre a ciò, MP3 ed altre codifiche applicano algoritmi che filtrano il segnale originale, eliminando suoni a loro volta coperti da suoni più forti e che il nostro orecchio non potrebbe correttamente percepire. Questo modulo introduce uno dei maggiori problemi, in quanto le radio utilizzano encoder con algoritmi lossy, ACC o MP3, con bassi livelli di bitrate, tipicamente 64 kbit/s o 128 kbit/s, che tagliano le frequenze sopra i 18 KHz.

Infine, una volta trasformato il segnale in digitale e compresso dall'encoder, questo viene mandato ad un server di streaming il quale lo rende disponibile, tramite una connessione TCP/IP, a tutti gli utenti che possiedono un player in grado di decodificare il segnale, che sia MP3, ACC o altro.

2.8 Stato dell'arte

Le stazioni AM/FM di trasmissione costituiscono una gran parte delle offerte audio online. Molti webcast offrono una programmazione simile alle loro stazioni terrestri. Alcuni utilizzano gli stessi spot pubblicitari on-line, ma molti scelgono di utilizzare la tecnologia di inserimento degli annunci che gli consente di monetizzare separatamente i loro stream. Prima di vedere queste tecnologie, nel paragrafo seguente vengono descritti i metodi e le tecnologie di profiling più utilizzate. Come visto ampiamente in questo capitolo, la personalizzazione del servizio è fondamentale per gli inserzionisti e per farla sono necessarie efficienti tecniche di profiling degli utenti.

2.8.1 Profilazione degli utenti

In questo mercato competitivo, i profili utente diventano molto importanti per i fornitori di servizi al fine di ottenere una personalizzazione del servizio che abbia successo. I servizi personalizzati mirano a soddisfare le esigenze e le preferenze degli utenti. Il successo di questi servizi si basa su come il fornitore di servizi conosca bene l'utente e quanto bene questo è riflesso sul servizio.

I profili utente sono la rappresentazione degli utenti e sono il risultato del processo di profilazione degli utenti. Ci sono due sfide principali nella profilazione degli utenti. Questi sono la generazione di un profilo utente iniziale per un nuovo utente e il continuo aggiornamento delle informazioni sul profilo per adattare la modifica delle preferenze, interessi e necessità dell'utente. In letteratura sono stati proposti due fondamentali metodi di profilazione per affrontare queste sfide. Questi metodi sono i "content-based", basati sui contenuti e i "collaborative methods", metodi di collaborazione. Entrambi questi metodi hanno delle limitazioni e al fine di superarle, è stata proposta la profilazione che è una combinazione di questi due metodi.

Un profilo utente è un insieme di informazioni che rappresenta un utente tramite regole, impostazioni, esigenze, interessi, comportamenti e preferenze [8]. Questa raccolta di informazioni personali può essere rappresentata come dati statici (ad esempio paese nativo) o dati dinamici (ad es. esigenze). Il contenuto e la quantità delle informazioni all'interno di un profilo utente può variare a seconda dell'area di applicazione. Tuttavia, indipendentemente dalle informazioni, l'accuratezza del profilo utente si basa sul modo in cui vengono raccolte e organizzate le informazioni utente e con quale precisione queste informazioni riflettono l'utente. In altre parole, questo dipende dal processo di profilazione e dal modo in cui le informazioni vengono raccolte, organizzate ed interpretate per creare la sintesi e la descrizione dell'utente [31]. Nella letteratura ci sono due modi fondamentali per recuperare le informazioni sull'utente. Queste sono chiamate informazioni esplicite o implicite. Nel metodo esplicito, le informazioni relative all'interesse dell'utente e le preferenze sono fornite esplicitamente dall'utente al sistema. Il lato negativo di questo metodo è che i profili espliciti hanno una natura statica e sono validi solo finché l'utente non cambia parametri di interesse e preferenze [16]. Nei metodi impliciti, invece, le informazioni sono catturate automaticamente tramite le interazioni che l'utente ha con il sistema; in questo modo, il profilo utente creato viene chiamato profilo dinamico.

La profilazione dinamica analizza i pattern di comportamento dell'utente (ad esempio la cronologia di utilizzo) per determinare gli interessi dell'utente. Qui, l'accuratezza del profilo utente dipende dalla quantità di dati generati attraverso l'interazione dell'utente con il sistema.

È anche possibile produrre un profilo utente ibrido che può essere costruito in due modi. Il primo modo inizia utilizzando l'esplicito in modo da raccogliere i dati iniziali, seguiti da quelli impliciti per aggiornare il profilo utente [28], oppure nel secondo modo, le tecniche implicite sono seguite dalle tecniche esplicite. In generale, i metodi ibridi sono più efficienti di quelli fondamentali, la tabella 2.1 mostra un confronto tra i tipi di profilo.

Due metodi fondamentali di profilazione utenti sono i "Content based" e i "Collaborative" [17]. È anche possibile utilizzare l'ibrido di questi due metodi. Le sezioni che seguono forniscono informazioni su ciascuno di questi metodi e sulle loro tecniche.

Tipo profilo utente	Descrizione	Tecniche utilizzate	Vantaggi	Svantaggi
Profilo utente esplicito	L'utente crea manualmente il profilo	Questionari e classifiche	Le informazioni raccolte sono di alta qualità	L'utente deve aggiornare continuamente le informazioni sul profilo
Profilo utente implicito	Il sistema genera il profilo usando le interazioni dell'utente con il sistema	Algoritmi di "machine learning"	Lo sforzo da parte dell'utente è minimo e gli aggiornamenti sono automatici	Inizialmente richiede tante interazioni tra l'utente e il sistema prima di creare un profilo accurato
Profilo utente ibrido	Combinazione dei due tipi	Tecniche implicite ed esplicite	Riduce i punti deboli e promuove i vantaggi delle due tecniche utilizzate	N/A

Tabella 2.1: Confronto tra i tipi di profilo

Metodi content-based

Il metodo basato sui contenuti, anche definito come filtraggio basato sui contenuti, assume che l'utente mostri lo stesso comportamento quando nelle stesse circostanze. Quindi in questo metodo, il comportamento corrente dell'utente è previsto dal suo comportamento passato, i profili utente sono rappresentati come query e il sistema seleziona gli elementi che hanno un'elevata correlazione al contenuto del profilo utente. La dipendenza dei contenuti è il principale inconveniente del filtraggio basato sui contenuti, pertanto, questo metodo si comporta male se il contenuto dell'articolo è molto limitato e non può essere analizzato facilmente dal filtraggio. Inoltre, gusti eclettici e scelte ad-hoc causano cattive prestazioni se ci si basa solo sulle scelte precedenti degli utenti. Per esempio, è possibile avviare un sistema che utilizzi un approccio basato sul contenuto raccomandando libri di storia a un professionista del computer che di solito acquista libri nel campo IT ma che per caso acquista una volta un libro di storia per suo fratello.

Metodi collaborative

Il metodo collaborativo, anche definito come filtraggio collaborativo, presuppone che gli utenti che appartengono allo stesso gruppo (ad es. stessa età, sesso o classe sociale) si comportino allo stesso modo e pertanto abbiano profili simili [8]. Il metodo collaborativo è basato sui modelli di rating di utenti simili. In questo metodo persone con modelli di rating simili, ovvero persone con simili gusti, sono chiamate "persone con la stessa mentalità". A differenza del metodo basato sui contenuti, il metodo

ignora i contenuti dell'elemento e fa la raccomandazione degli articoli solo sulla base delle valutazioni degli elementi degli utenti simili [17]. Ci sono due principali inconvenienti in questo metodo, la sparsità e il problema del primo rater. La sparsità è la situazione in cui c'è una mancanza di valutazioni disponibili, causato da un numero insufficiente di utenti o da poche valutazioni per utente. Inoltre, il problema del primo rater, noto anche come problema di "cold-start", può essere osservato quando un nuovo utente ha un basso numero di valutazioni.

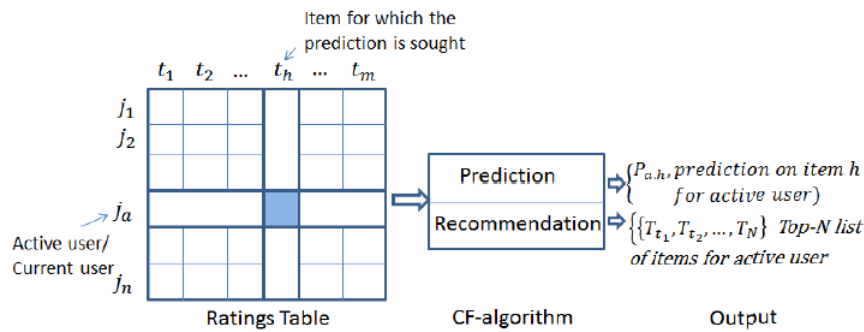


Figura 2.7: Principio base del metodo collaborativo

Metodi ibridi

Il metodo ibrido, chiamato anche metodo di filtraggio ibrido utilizza entrambi i metodi per combinarne i vantaggi e superare i limiti di entrambi i metodi. Questo metodo garantisce l'immediata disponibilità di un profilo per ogni utente. Il sistema che impiega il metodo ibrido fornisce una descrizione più accurata degli interessi e delle preferenze dell'utente, in quanto controlla e recupera continuamente le informazioni relative all'utente attraverso l'interazione del sistema utente [8]. In genere, il metodo ibrido assegna al nuovo utente un profilo predefinito con l'utilizzo del metodo collaborativo e migliora ulteriormente il profilo utilizzando il metodo basato sui contenuti. Nella letteratura esistono quattro tecniche ibride per profilare gli utenti [28]. Questi sono chiamati "static content profiling", "dynamic content profiling", "static collaborative profiling", e "dynamic collaborative profiling". la tecnica "static content profiling" è la combinazione di "static profiling" e del metodo basato sui contenuti. Qui vengono raccolte le informazioni sugli interessi degli utenti durante la registrazione. Nel "dynamic content profiling", le informazioni e gli interessi degli utenti vengono recuperati tramite il monitoraggio del comportamento dell'utente. Nello "static collaborative profiling", invece, le informazioni relative agli interessi dell'utente vengono raccolte in base alle richieste esplicite dell'utente. Qui il raggruppamento degli utenti è fatto esplicitamente. Infine, nel dynamic collaborative profiling, la raccolta delle informazioni e il raggruppamento degli utenti vengono effettuate in base a le risposte dinamiche degli utenti con comportamenti simili. La tabella 2.3 confronta i suddetti tre profili utente principali.

Metodo di profilazione	Descrizione	Tecniche utilizzate	Vantaggi	Svantaggi
Content-based	Filtraggio del contenuto da un flusso di dati basato sull'estrazione di caratteristiche che sono state espresse	Vector Space model, Latentsemantic indexing, Learninginformation agents, Neuralnetwork agents	Analisi obiettiva da grandi fonti di materiale digitale (ad esempio, multimediali) senza un grande coinvolgimento degli utenti	Dipendenza dal contenuto, difficile da introdurre valutazioni inaspettate
Collaborative	Filtraggio degli elementi basato sulle somiglianze tra il profilo utente di destinazione e il gruppo di appartenenza	Memory-based e Model-based	Lo sforzo da parte dell'utente è minimo e gli aggiornamenti sono automatici	Sparsità: scarse capacità di predizione quando un nuovo elemento viene introdotto nel database a causa della mancanza di valutazioni. First-rated: scarse valutazioni fatte dai nuovi utenti finché non hanno abbastanza voti nei loro profili per un confronto accurato con altri utenti
Ibrido	Combinazione dei due tipi	Entrambe	Riduce i punti deboli e promuove i vantaggi delle due tecniche utilizzate	I punti deboli possono eliminare i punti forti se l'ibrido è creato ingenuamente

Tabella 2.3: Confronto tra i vari metodi

2.8.2 Tecnologie d'inserimento

Le aziende di tecnologia che facilitano l'inserimento di annunci pubblicitari includono, ma non sono solo:

- **Triton Digital:** La piattaforma di annunci Triton (Tap) è una piattaforma

tecnologica che fornisce annunci, gestisce campagne, compresa la separazione concorrenziale degli annunci e traccia banner di accompagnamento che talvolta passano con gli annunci audio; Tap OnDemand è personalizzato per i servizi di radio Internet, mentre Tap Live è progettato per le emittenti AM / FM.

- **AdsWizz:** AdsWizz è un prodotto di inserzione chiamata Audio Injection for Streamers (AIS). È uno strumento agnostico per la piattaforma (flusso di trasmissione FM o radio in Internet) che funziona con molti formati audio nei flussi di musica e può funzionare con diversi suggerimenti di iniezione, tra cui metadati, bip e jingle che precedono le interruzioni di annuncio.
- **Abacast:** Abacast è fornitore di servizi di streaming per la radio. La soluzione per l'inserimento di annunci è basata su cloud e offre una flessibilità di targeting personalizzata, ad esempio iniettando un annuncio agli ascoltatori di smartphone e un altro ad ascoltatori maschi di età tra i 25-34.

L'audio in streaming è intrinsecamente mobile, su smartphone, tablet e dashboard auto. Al di là di questo vantaggio fondamentale, le tecnologie mobili specifiche portano ulteriore efficacia agli inserzionisti in streaming.

Una tale tecnologia è geo-fencing, che consiste in un perimetro virtuale di una posizione reale. Essa sfrutta i segnali GPS, torri per telefoni cellulari e WiFi per individuare singoli ascoltatori e controllare la loro ricezione dei contenuti. Geo-fencing è utilizzato per pubblicare annunci a consumatori che sono all'interno del perimetro specificato come geolocalizzazione da parte dell'inserzionista. Così, ad esempio, una concessionaria locale di auto potrebbe pubblicizzare una vendita a fine anno ai consumatori che ascoltano il flusso in una distanza di guida dello showroom. La consegna precisa di campagne geo-situate è una caratteristica digitale che separa notevolmente l'audio streaming dalla trasmissione. Un'altra tecnologia, che si ritrova dietro la geo-fencing nello sviluppo e nell'adozione, ma più futuristica nell'attuazione, si chiama audio faro.

I beacon audio sono suoni ad ultrasuoni che innescano contenuti speciali da utilizzare nelle applicazioni audio mobili. Esse richiedono alcune parti in essere - apparecchi di invio, un dispositivo ricevente e applicazioni che hanno il riconoscimento del beacon incorporato in essi e possono fornire il contenuto speciale. I beacon audio non sono udibili alle persone, ma sono udibili anche per le applicazioni che li ricevono. Un esempio di marketing del beacon potrebbe avvenire in uno stadio sportivo, dove un beacon di trasmissione attiva applicazioni abilitate tra gli spettatori, indirizzandoli ad annunci pertinenti. Un articolo del Wall Street Journal[12] ha descritto la commercializzazione del beacon condotta da Sony Music, che ha utilizzato la tecnologia del beacon fornita dalla tech startup Lisnr per attivare contenuti speciali streaming per i consumatori a piedi a Times Square a New York City. A questi consumatori è stata offerta la possibilità di visualizzare in anteprima il prossimo album dell'artista J. Cole, generando pre-ordini di album e social buzz.²

²Il buzz marketing, detto anche marketing non convenzionale è quell'insieme di operazioni di marketing non convenzionale volte ad aumentare il numero e il volume delle conversazioni riguardanti un prodotto o un servizio e, conseguentemente, ad accrescerne la notorietà e la buona reputazione.

2.9 Motivazioni

La radio terrestre partecipa alla rivoluzione dello streaming mobile distribuendo i flussi di webcast in applicazioni mobili, contemporaneamente, l'industria per lo sviluppo del software ha molte aziende che si specializzano nella creazione di applicazioni mobili per singoli brand.

L'audio in streaming, che fa parte dell'ampio movimento della tecnologia digitale, è sempre in evoluzione. Il cambiamento avviene sia dal lato della programmazione e distribuzione dei contenuti ai consumatori sia dalla parte del marketing in flussi audio e monetizzazione delle piattaforme di streaming.

Come abbiamo visto in questo capitolo, la internet radio è in continua espansione e viene ascoltata principalmente da dispositivi mobile; per le radio la maggior fonte di guadagno è quella delle inserzioni pubblicitarie, dall'altra parte, gli inserzionisti richiedono servizi pubblicitari sempre più personalizzati.

Mentre nella radio terrestre la personalizzazione avviene sulle fasce orarie e/o sulla posizione demografica, come visto nel paragrafo 2.2, nella internet radio, la personalizzazione può essere ampliata.

La tecnologia per la personalizzazione è ben definita (vedi par. 2.8.1), quello che invece non è ancora perfettamente delineato è come, dopo aver profilato un utente, fornire un'inserzione adatta a quel profilo.

A questo scopo nel par 2.8.2 sono state descritte alcune tecniche utilizzate; quello che questo documento si promuove di fare è di creare un'alternativa a quello che al momento, lo stato dell'arte propone, tutto questo mantenendo l'infrastruttura già presente in modo da rendere l'implementazione semplice ed economica.

Capitolo 3

Processamento del segnale audio

3.1 Musica e fisica

Un suono è una vibrazione che si propaga attraverso l'aria (o l'acqua) e può essere "decriptata" dalle orecchie. Ad esempio, quando si ascolta un mp3, gli auricolari producono vibrazioni che si propagano nell'aria fino a raggiungere le orecchie. La luce è anche una vibrazione ma non si può sentire perché le orecchie non possono "decriptarla" (ma gli occhi si).

Una vibrazione può essere modellata da forme d'onda sinusoidali. In questa sezione vedremo come la musica può essere descritta fisicamente e tecnicamente.

3.1.1 Toni puri e suoni reali

Un tono puro è un tono con una forma d'onda sinusoidale. Esso è caratterizzata da:

- **La sua frequenza:** il numero di cicli al secondo, la sua unità di misura è l'Hertz (Hz), ad esempio $100\text{Hz} = 100$ cicli al secondo.
- **La sua ampiezza:** la dimensione di ciascun ciclo.

Queste caratteristiche vengono decriptate dall'orecchio umano per formare un suono. L'umano può sentire toni puri da 20 Hz a 20000 Hz (per le orecchie migliori) e questa gamma diminuisce con l'età. In confronto, la luce che si vede è composta da onde sinusoidali da $4 \cdot 10^{14}\text{Hz}$ a $7.9 \cdot 10^{14}\text{Hz}$.

La percezione umana dell'intensità di un tono puro dipende dalla sua frequenza, ad esempio, un tono puro ad ampiezza 10 della frequenza di 30Hz sarà più silenzioso di un tono puro all'ampiezza 10 della frequenza di 1000Hz.

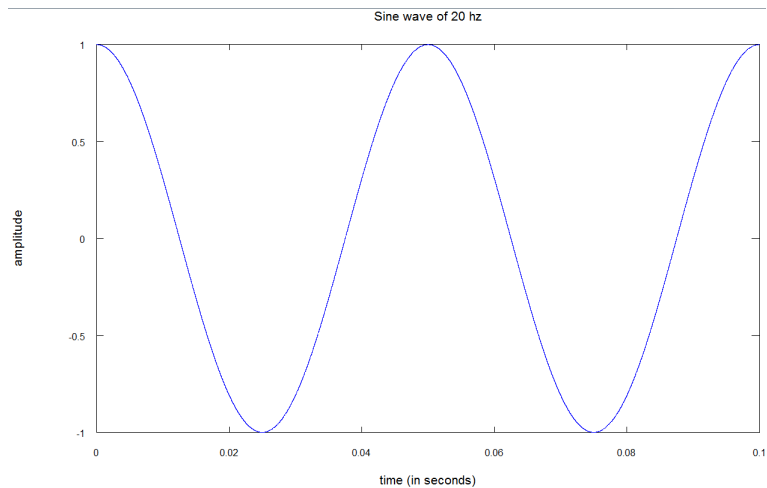


Figura 3.1: Rappresentazione di un'onda sinusoidale pura di frequenza 20hz e di ampiezza 1

I toni puri non esistono naturalmente, ma ogni suono del mondo è la somma di toni puri multipli a diverse ampiezze.

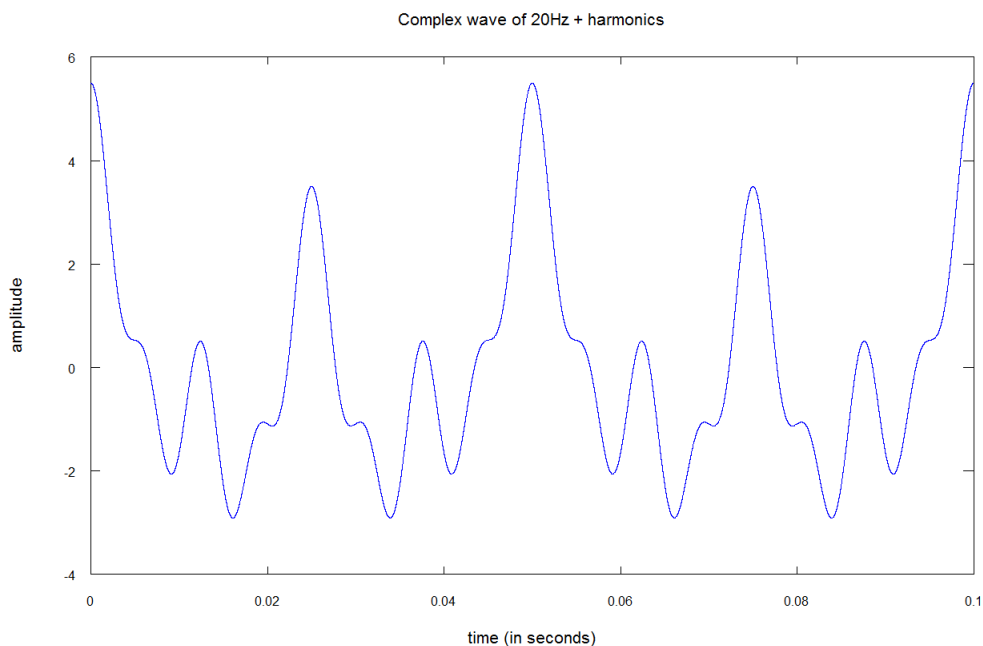


Figura 3.2: Rappresentazione di una composizione di onde sinusoidali

In questa figura si può vedere la rappresentazione di un suono più realistico che è la composizione di più onde sinusoidali:

- Un'onda sinusoidale pura di frequenza 20hz e ampiezza 1
- Un'onda sinusoidale pura di frequenza 40hz e ampiezza 2
- Un'onda sinusoidale pura di frequenza 80hz e ampiezza 1.5
- Un'onda sinusoidale pura di frequenza 160hz e ampiezza 1

Un suono reale può essere composto da migliaia di toni puri.

3.2 Digitalizzazione

Quando si ascolta la musica, a meno che non si utilizzi un vinile, si usa un file digitale (mp3, acc , ogg, audio CD, ecc.). Quando gli artisti producono musica, essa però è in analogico. La musica viene digitalizzata per essere memorizzata e riprodotta da dispositivi elettronici (come computer, telefoni, lettori mp3, cd player). In questa parte vedremo come passare da un suono analogico a quello digitale, sapere come la musica digitale è fatta ci aiuterà ad analizzarla e manipolarla nelle prossime parti.

3.2.1 Campionamento

I segnali analogici sono segnali continui, il che significa che se si prende un secondo di un segnale analogico, è possibile dividere questo secondo in più parti che durano una frazione di secondo. Nel mondo digitale, non ci si può permettere di memorizzare una quantità infinita di informazioni, è necessario quindi disporre di un'unità minima, ad esempio 1 millisecondo. Durante questa unità di tempo il suono non può cambiare, quindi questa unità deve essere abbastanza breve in modo che il segnale digitale suoni come quello analogico e abbastanza grande per limitare lo spazio necessario per la memorizzazione.

In parole povere il campionamento consiste nell'andare a registrare il valore del segnale analogico in diversi istanti di tempo o dello spazio. Il tempo T che intercorre tra una valutazione e l'altra si chiama intervallo spaziale o temporale di campionamento. La frequenza di campionamento $f_c = \frac{1}{T}$ è invece il reciproco dell'intervallo spaziale o temporale di campionamento.

L'unità standard di tempo nella musica digitale è di 44 100 unità (o campioni) al secondo. Ma da dove viene questo 44,1 kHz?

Il teorema che stabilisce quale sia la frequenza minima di campionamento con una determinata caratterizzazione in frequenza (trasformata di Fourier) affinché il segnale analogico possa essere ricostruito a valle a partire da quello discreto in input è il teorema del campionamento di Shannon-Nyquist, ovvero: $f_c > 2 \cdot f_m$

Dove f_c è la frequenza di campionamento e f_m è la massima frequenza dello spettro del segnale da campionare.

L'idea principale è che un segnale d'onda sinusoidale a una frequenza F abbia bisogno di almeno 2 punti per ciclo per poter essere identificato. Se la frequenza del campionamento è almeno doppia rispetto alla frequenza del segnale si avranno almeno 2 punti per ciclo.

Cerchiamo di capire con un'immagine un esempio di buon campionamento. Nella figura 3.3 è rappresentato un suono a 20Hz che viene digitalizzato utilizzando una frequenza di campionamento a 40Hz:

- La curva blu rappresenta il suono a 20 Hz
- Le croci rosse rappresentano il suono campionato
- La linea verde è l'interpolazione del suono campionato.

Anche se non ha la stessa forma, né la stessa ampiezza, la frequenza del segnale campionato resta la stessa.

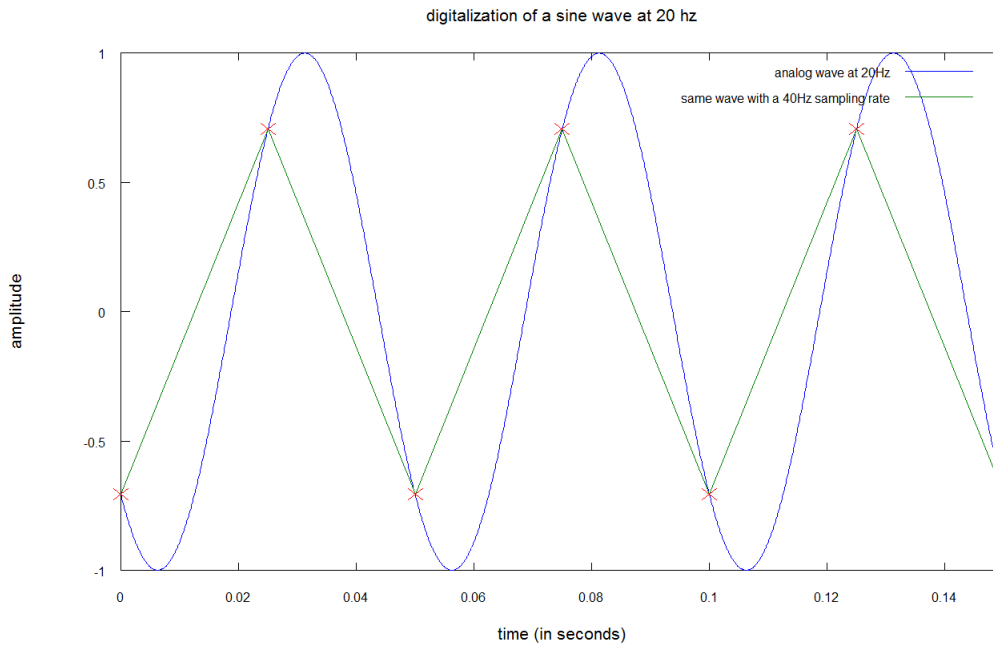


Figura 3.3: Rappresentazione di un onda sinusoidale a 20 hz campionata a 40hz

Di seguito vediamo invece un esempio di campionamento fatto male. Nella figura 3.4, un suono a 20 Hz viene digitalizzato con una frequenza di campionamento a 30Hz. Questa volta la frequenza del segnale campionato non è la stessa del segnale originale. Se si guarda attentamente, si può vedere che un ciclo nel segnale campionato rappresenta due cicli nel segnale originale. Questo è un caso di "under sampling" (sotto campionamento).

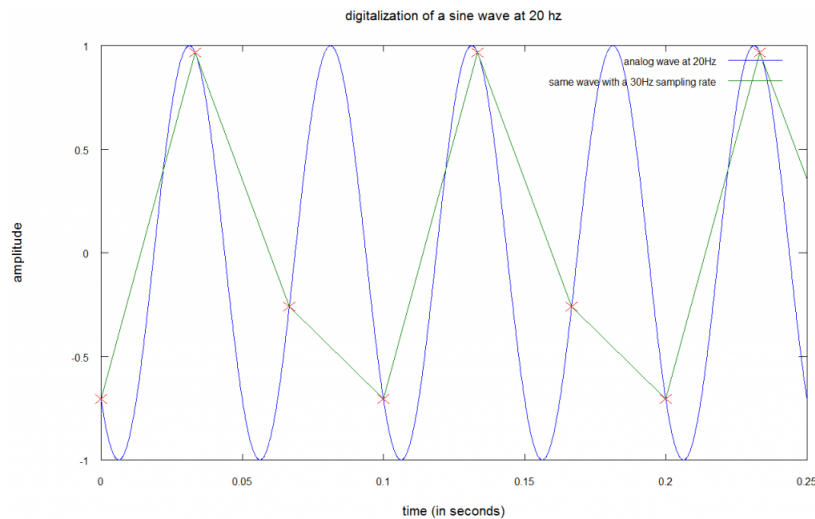


Figura 3.4: Rappresentazione di un onda sinusoidale a 20 hz campionata a 30hz

Questo caso mostra anche il fenomeno dell'**aliasing**. Se si desidera digitalizzare un segnale compreso tra 0Hz e 20kHz, è necessario rimuovere dal segnale le frequenze al di sopra dei 20kHz prima del campionamento, altrimenti queste frequenze saranno trasformate in frequenze tra 0Hz e 20Khz e quindi verranno aggiunti suoni indesiderati.

3.2.2 Quantizzazione

Abbiamo visto come digitalizzare le frequenze di una musica analogica, ma che dire dell'intensità della musica? L'intensità è una misura relativa, essa misura la variazione tra il livello più basso e quello più alto del suono all'interno di una canzone. Quando si misura una grandezza fisica, l'insieme di valori che essa può assumere in natura è un insieme continuo e composto da infiniti punti. A causa dell'impossibilità di una rappresentazione reale a precisione infinita, nelle comunicazioni di tipo numerico o digitale, il valore della grandezza in questione deve essere convertito in formato discreto. Affinché una grandezza sia trasmissibile e codificabile con un numero finito di bit ovvero in forma numerica, è però necessario che essa possa assumere solo un numero finito di valori di codominio discreti; ciò avviene tramite un successivo processo di quantizzazione del valore in ordinata della grandezza in questione.

Per ottenere ciò, i valori possibili della grandezza in questione vengono innanzitutto limitati tra un massimo ed un minimo intorno a dei valori discreti preventivamente definiti definendo così le relative regioni di decisione e la dinamica del quantizzatore stesso: in tal modo il valore analogico della grandezza originaria, in corrispondenza del valore campionato in ascissa, verrà ricondotto al più prossimo dei valori discreti preventivamente definiti tramite il processo di decisione.

La figura 3.5 è un esempio di una bassa quantizzazione di un segnale audio:

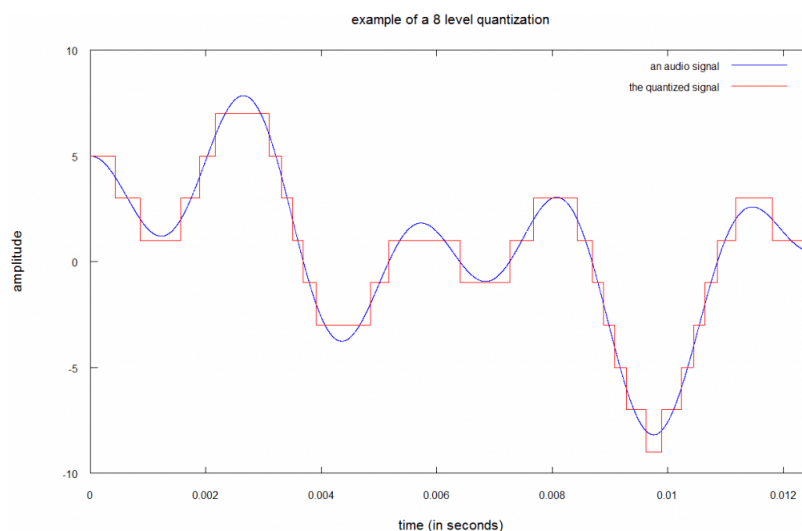


Figura 3.5: Rappresentazione di un segnale quantizzato a 8 livelli

Come si può vedere, il suono risultante (in rosso) è molto alterato. La differenza tra il suono reale e quella quantizzata viene chiamata errore di quantizzazione o rumore di quantizzazione. Questa quantizzazione a 8 livelli è anche definita una quantizzazione a 3 bit perché necessita solo di 3 bit per implementare gli 8 diversi livelli $8 = 2^3$.

Quello in figura 3.6 è lo stesso segnale con una quantizzazione a 64 livelli (o quantizzazione a 6 bit).

Anche se il suono risultante è ancora alterato, è comunque più simile al suono originale. Fortunatamente, gli esseri umani non hanno orecchie sensibili. La quantizzazione standard è codificata su 16 bit, il che significa 65536 livelli. Con questa quantizzazione, il rumore di quantizzazione risulta essere abbastanza basso.

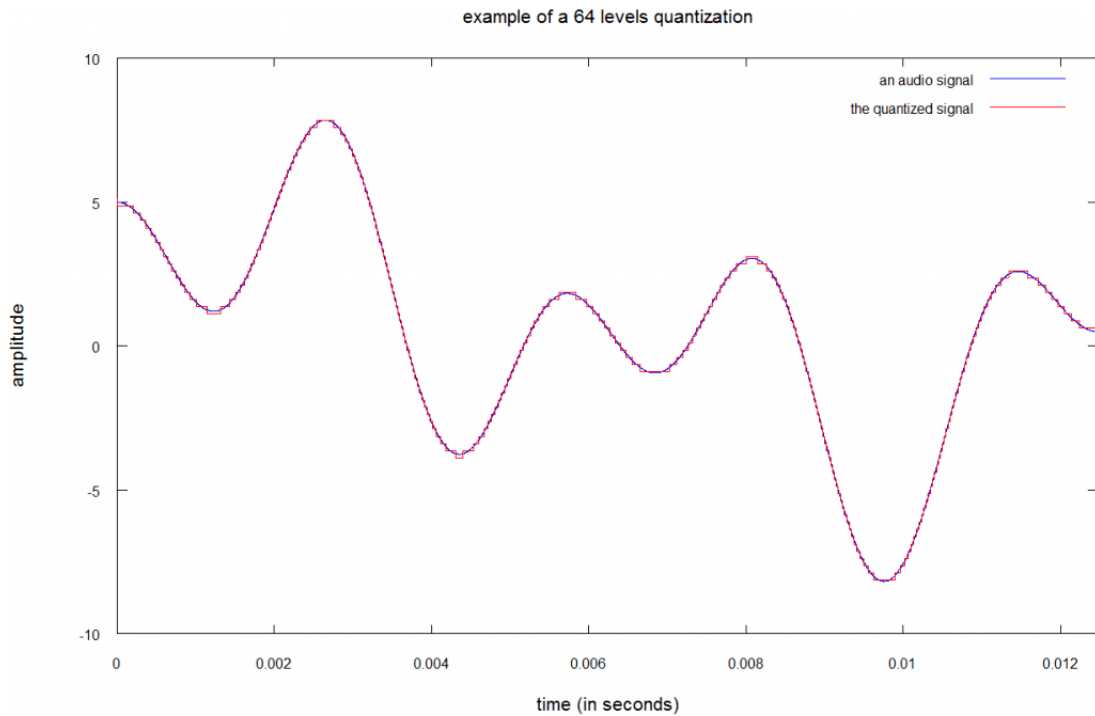


Figura 3.6: Rappresentazione di un segnale quantizzato a 64 livelli

3.2.3 Pulse Coded Modulation

La modulazione PCM (Pulse Coded Modulation) è uno standard che rappresenta i segnali digitali. Viene utilizzato dai CD e dalla maggior parte dei dispositivi elettronici. Ad esempio, quando si ascolta un file mp3 nel computer, telefono o tablet, l'mp3 viene automaticamente trasformato in un segnale PCM e quindi inviato alle cuffie.

Un flusso PCM è un flusso di bit organizzati. Può essere composto da più canali, ad esempio, una musica stereo ha 2 canali. In un flusso, l'ampiezza del segnale è divisa in campioni; il numero di campioni al secondo corrisponde alla frequenza di campionamento della musica, una musica campionata a 44,1 kHz avrà quindi 44100 campioni al secondo. Ogni campione fornisce l'ampiezza (quantizzata) del suono nella frazione corrispondente di secondi. Esistono più formati PCM, ma il più utilizzato è il formato (lineare) PCM 44,1 kHz, a 16 bit stereo. Questo formato ha 44 100 campioni per ogni secondo di musica. Ogni campione richiede 4 byte:

- 2 byte (16 bit) per l'intensità dell'altoparlante sinistro (da -32.768 a 32.767)
- 2 byte (16 bit) per l'intensità dell'altoparlante destro (da -32.768 a 32.767)

3.3 Dal suono digitale alle frequenze

Abbiamo visto come passare da un suono analogico a quello digitale. Ma come si possono ottenere le frequenze all'interno di un segnale digitale?

Per i segnali analogici (e quindi continui), vi è una trasformazione chiamata trasformazione di Fourier continua. Questa funzione trasforma una funzione del tempo in una funzione delle frequenze. In altre parole, se si applica la trasformata di Fourier

PCM 16-bit depth stereo sample

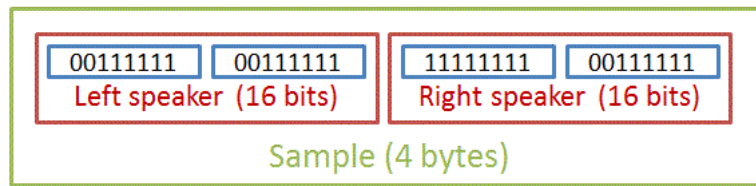


Figura 3.7: Rappresentazione binaria di un campione PCM 16 bit stereo

su un suono, otterremo le frequenze (e le loro intensità) all'interno ma ci sono due problemi: i suoni digitali sono suoni finiti e quindi non continui e per avere una migliore conoscenza delle frequenze all'interno di una musica, dobbiamo applicare la trasformazione di Fourier su piccole parti del segnale audio, in modo da sapere quali sono le frequenze per ogni piccola parte di una traccia audio. Fortunatamente, c'è un'altra funzione matematica, la trasformata di Fourier discreta (DFT), che funziona con alcune limitazioni.

3.3.1 Trasformata discreta di Fourier

La DFT (Discrete Fourier Transform)[22] si applica a segnali discreti e fornisce uno spettro discreto cioè le frequenze all'interno del segnale. La seguente è la formula per trasformare un segnale digitale in frequenze.

$$X_n = \sum_{k=0}^{N-1} x_k e^{-jk \frac{2\pi}{N} n} \quad k = 0, \dots, N - 1$$

dove:

- N è la dimensione della finestra: il numero di campioni che compongono il segnale
- X_n rappresenta l' n -esimo bin¹ di frequenze
- X_k è il k -esimo campione del segnale audio

Ad esempio, se si ha un segnale con una finestra di 4096 campioni, la precedente formula va applicata 4096 volte.

Come si è potuto notare, si è parlato di bin di frequenze e non di frequenza. Il motivo è che la DFT dà uno spettro discreto, un bin di frequenze è l'unità di frequenza più piccola che la DFT può calcolare. La dimensione del bin (chiamata spettro di risoluzione o risoluzione di frequenza) equivale alla frequenza di campionamento del segnale diviso per la dimensione della finestra (N). Nel nostro esempio, con una finestra di campionamento di 4096 campioni e una frequenza standard di campionamento audio a 44,1 kHz, la risoluzione di frequenza è di 10,77 Hz (ad eccezione del 0°bin che è speciale):

- Il 0°bin rappresenta le frequenze tra 0Hz e 5.38Hz
- Il 1°bin rappresenta le frequenze tra 5.38Hz e 16.15Hz

¹bin: intervallo di frequenze

- Il 2°bin rappresenta le frequenze tra 16,15 e 26,92 Hz
- Il 3°bin rappresenta le frequenze tra 26,92 Hz e 37,68 Hz
- ...

Ciò significa che la DFT non può dissociare 2 frequenze che sono vicine a 10.77Hz. Ad esempio le note a 27Hz, 32Hz e 37Hz finiscono nello stesso bin. Se la nota a 37Hz è molto intensa si saprà solo che il terzo bin ha un'intensità alta. Questo diventa problematico quando bisogna dissociare le note nelle ottave più basse.

È possibile migliorare la risoluzione delle frequenze aumentando la dimensione della finestra ma ciò significa perdere le frequenze/note che cambiano velocemente all'interno della musica. Aumentare la finestra significa prendere più campioni e quindi aumentare il tempo per prendere la finestra, ad esempio, con 4096 campioni, la durata della finestra è di 0,1 secondi e la risoluzione di frequenza è di 10,7 Hz, quindi è possibile rilevare una variazione ogni 0,1 secondi, con 16384 campioni, la durata della finestra è 0,37 secondi e la risoluzione di frequenza è 2,7 Hz quindi è possibile rilevare una variazione ogni 0,37 secondi.

Non ci sono finestre "migliori" o "peggiori". Ogni finestra ha le sue caratteristiche e dipende dal problema che si desidera risolvere utilizzarne un tipo o un altro. Una finestra rettangolare ha caratteristiche di risoluzione eccellenti per sinusoidi di resistenza paragonabile, ma è una scelta scadente per sinusoidi di ampiezze disparate (come nel caso di una canzone perché le note musicali non hanno la stessa intensità) mentre la finestra di Blackman è meglio evitarla nel caso in cui la perdita di spettro di frequenze frequenti nasconde frequenze deboli.

3.3.2 Fast Fourier Transform

Se si esamina nuovamente la formula 3.3.1 della DFT, si può vedere che per calcolare un bin si devono fare N addizioni e N moltiplicazioni (dove N è la dimensione della finestra), quindi, ottenere N bin richiede $2 \cdot N^2$ operazioni, che sono molte.

Per esempio, supponiamo di avere una canzone di tre minuti a 44,1 kHz e di calcolare lo spettrogramma della canzone con una finestra di 4096 campioni, in questo modo si dovrebbero calcolare 10.7 (44100/4096) DFT al secondo e 1938 DFT per la canzone completa.

Ogni DFT ha bisogno di $3.35 \cdot 10^7$ operazioni e quindi per ottenere lo spettrogramma della canzone bisogna fare $6,5 \cdot 10^{10}$ operazioni.

Supponiamo di avere una raccolta musicale di mille canzoni di tre minuti, avremmo bisogno di $6,5 \cdot 10^{13}$ operazioni per ottenere gli spettrogrammi canzoni. Anche con un buon processore, ci vorrebbero giorni/mesi per ottenere il risultato.

Fortunatamente, ci sono implementazioni più veloci della DFT chiamate FFT (Fast Fourier Transforms) [33]. Alcune implementazioni richiedono solo operazioni di $1.5 \cdot N \cdot \log(N)$ operazioni. Per la stessa collezione di musica, utilizzando la FFT anziché la DFT ci vorrebbero 340 addizioni in meno ($1.43 \cdot 10^{11}$) e quindi ci vorrebbero soltanto minuti/ore per ottenere il risultato.

Questo esempio mostra un altro compromesso: anche se aumentando la dimensione della finestra migliora la risoluzione di frequenza, aumenta anche il tempo di calcolo. Per la stessa collezione di musica, se si calcola lo spettrogramma usando una finestra di campionamento da 512 (risoluzione di frequenza di 86 Hz), con la

FFT si ottiene il risultato con $1,07 \cdot 10^{11}$ operazioni, circa 1/4 di tempo più veloce che con una finestra di 4096 campioni (risoluzione di frequenza di 10,77 Hz).

Gli algoritmi della Discrete Fourier Transform (DFT), e le versioni collegate della Fast Fourier Transform (FFT), sono normalmente utilizzati per il calcolo di ampiezza e fase dell'intero spettro di un segnale. Generalmente, tuttavia, non è necessario conoscere l'intero spettro del segnale, bensì è sufficiente la conoscenza di una singola armonica, o di un esiguo numero di armoniche. I classici algoritmi di DFT e FFT non sono efficienti, dal punto di vista computazionale, per la misura di un singolo tono in frequenza, senza contare il notevole carico computazionale richiesto per il calcolo completo dello spettro del segnale.

La teoria dell'elaborazione numerica dei segnali, tuttavia, offre un'alternativa qualora sia necessario misurare solo un'armonica e sia richiesta una bassa complessità computazionale. L'algoritmo è noto con il nome di algoritmo di Goertzel, il cui dettaglio è riportato nel paragrafo 3.4.

3.4 Algoritmo di Goertzel

L'algoritmo di Goertzel è molto vantaggioso rispetto alla FFT principalmente per la bassa complessità computazionale nel calcolo di una singola armonica. Un altro vantaggio dell'algoritmo di Goertzel è la sua natura incrementale, che consente il calcolo "distribuito" nel tempo dell'armonica di interesse durante l'esecuzione dell'algoritmo. Nella FFT invece è richiesta l'intera finestra di acquisizione del segnale prima di procedere con i calcoli.

Questo algoritmo è un'effettiva forma convoluzionale della trasformata discreta di Fourier per il calcolo diretto del valore di Fourier in corrispondenza di una frequenza selezionata, vale a dire, valutare solo un bin selezionato dello spettro di Fourier. L'algoritmo è utilizzato principalmente come rivelatore di DTMF (Dual Tone Multi-Frequency). Tuttavia, le sue caratteristiche permettono di utilizzarlo come filtro digitale [13].

L'algoritmo di Goertzel di base fornisce le componenti reali e immaginarie di frequenza così come una normale trasformata di Fourier discreta (DFT) o FFT e da esse è possibile calcolare la magnitudine e la fase. Esiste anche una versione ottimizzata dell'algoritmo che è ancora più veloce (e più semplice) rispetto al Goertzel di base, ma non dà le componenti reali e immaginari della frequenza, esso fornisce soltanto la magnitudine relativa quadrata dalla quale è possibile ottenere la magnitudine relativa (se necessario), ma non c'è modo di ottenere la fase.

3.4.1 Funzionamento dell'algoritmo base

In questo breve paragrafo, non si vogliono descrivere le basi teoriche dell'algoritmo, per la quale si rimanda alla bibliografia [36], ma il suo funzionamento algoritmico.

Come per la FFT, si lavora con blocchi di campioni. Tuttavia, questo non significa che bisogna elaborare i dati in blocchi. L'elaborazione numerica è abbastanza breve da eseguire, utilizzando buffers di campioni, è possibile andare avanti ed elaborarli un set alla volta.

Prima di poter elaborare il Goertzel effettivo, è necessario eseguire alcuni calcoli preliminari:

1. Decidere la frequenza di campionamento.
2. Scegliere la dimensione del blocco, N.
3. Pre-elaborare coseni e seni
4. Pre-elaborare un coefficiente

Questi possono essere tutti pre-elaborati prima e poi codificati nel programma, salvando spazio in RAM, oppure è possibile calcolarli al volo.

La frequenza di campionamento è determinata dall'uso dell'applicazione. Ad esempio, nelle telecomunicazioni, è comune utilizzare una frequenza di campionamento di 8 kHz (8000 campioni al secondo). Quando si sceglie la frequenza di campionamento, si applicano le solite regole di Nyquist: la frequenza di campionamento dovrà essere almeno due volte la frequenza di interesse più alta.

La dimensione del blocco N è come il numero di punti in un equivalente FFT. Essa controlla la risoluzione di frequenza (chiamata anche larghezza del bin). Ad esempio, se la frequenza di campionamento è di 8 kHz e N è di 100 campioni, allora la larghezza del bin sarà di 80Hz. Questo porterebbe a rendere N il più alto possibile, per ottenere la massima risoluzione di frequenza. Il problema, come descritto precedentemente nel paragrafo 3.3.1, è che più è alto N, più ci vorrà per rilevare ogni tono; questo perché bisognerebbe aspettare più a lungo affinché si abbiano tutti i campioni.

Ad esempio, con un campionamento a 8 kHz, ci vorranno 100 ms per accumulare 800 campioni, se si sta cercando di rilevare toni di breve durata, bisognerà utilizzare valori compatibili di N. Il terzo fattore che influenza la scelta di N è la relazione tra la frequenza di campionamento e il target di frequenze; idealmente si desidera che le frequenze siano concentrate nei rispettivi bin. In altre parole, si vuole che le frequenze siano un multiplo intero della frequenza di campionamento diviso la dimensione del blocco. La buona notizia è che, a differenza della FFT, N non deve essere un potenza di due.

Una volta scelta la frequenza di campionamento e la dimensione del blocco N, bisogna fare alcuni pre-calcoli che verranno utilizzati durante il processamento:

Dati:

- N dimensione del blocco
- **sample_rate** frequenza di campionamento
- **target_freq** frequenza target
- $k = (int) \cdot (0.5 + \frac{N \cdot target_freq}{sample_rate})$

Si eseguono i seguenti pre-calcoli:

- $w = \frac{2 \cdot \pi}{N} \cdot k$
- *coseno* = $\cos w$
- *seno* = $\sin w$

- $coeff = 2 \cdot coseno$

Una volta effettuati i pre-calcoli, si passa all'elaborazione per campione; c'è bisogno di tre variabile chiamate **Q0**, **Q1** e **Q2**, **Q1** è il valore di **Q0** nell'ultimo step e **Q2** è il valore di **Q0** nel penultimo step; **Q1** e **Q2** devono essere inizializzati a zero all'inizio di ogni blocco di campioni e per ogni campione bisogna effettuare le tre seguenti equazioni:

- $Q0 = coeff \cdot Q1 - Q2 + sample$
- $Q2 = Q1$
- $Q1 = Q0$

Dopo aver eseguito le equazioni per campione N volte, bisogna vedere se la frequenza target è presente o no, a questo scopo si calcola la parte reale, la parte immaginaria e la magnitudine come segue:

- $real = (Q1 - Q2 \cdot coseno)$
- $imag = (Q2 \cdot seno)$
- $magnitude^2 = real^2 + imag^2$

Un semplice test sulla soglia della magnitudine ci dirà se la frequenza è presente o meno; prima di passare al prossimo blocco si resettano Q2 e Q1.

Capitolo 4

Metodologia

In questo capitolo vengono esposti alcuni concetti e tecnologie che poi verranno utilizzati nel prossimo capitolo per lo sviluppo dei due sistemi di segnalazione.

4.1 Sistemi di comunicazione

In generale un sistema di comunicazione ha lo scopo di trasferire informazione da un punto, detto sorgente o emettitore, in una certa posizione e ad un certo istante, ad un altro punto, detto ricevitore.

A seconda del numero e della disposizione degli utenti, i sistemi di comunicazione possono essere suddivisi in

- sistemi broadcast: sistemi che trasferiscono la medesima informazione a più utenti;
- sistemi punto–punto: sistemi che collegano due punti fissi nello spazio;
- sistemi portatili: sistemi di comunicazione trasportabili, ma che devono essere fissi al momento dell'utilizzo;
- sistemi mobili: sistemi in cui l'utente può essere in movimento durante la comunicazione.

Mantenendo un alto livello di generalità, possiamo scindere un sistema di comunicazione nei blocchi fondamentali rappresentati in Fig. 4.1.

In generale il messaggio informativo prodotto dalla sorgente può essere non elettrico. In ingresso dovremo avere perciò un trasduttore (ad esempio un microfono) che converta il messaggio in un segnale elettrico, detto segnale informativo. Analogamente, alla fine della catena, può essere presente un opportuno trasduttore (il trasduttore di uscita) per convertire il segnale elettrico ricevuto in una forma appropriata alla particolare applicazione (ad esempio un altoparlante). Una suddivisione fondamentale dei sistemi di comunicazione si ha in base al tipo di informazione che deve essere trasferita. In particolare distinguiamo:

- sorgenti di informazione analogiche: producono un segnale informativo definito su un insieme continuo, ossia un segnale che può assumere un qualsiasi valore tra un valore minimo ed un valore massimo. Un esempio di sorgente analogica è il microfono;

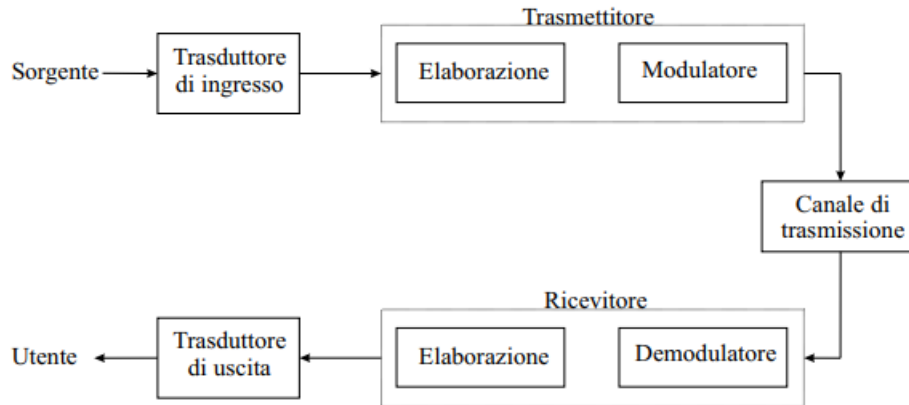


Figura 4.1: Blocchi fondamentali costituenti un sistema di comunicazione

- sorgenti di informazione discreta: producono un segnale informativo definito su un insieme con un numero finito di elementi, ossia un segnale che in ogni istante può essere uguale ad uno tra un certo numero di simboli o valori. Un esempio di sorgente discreta è la macchina da scrivere, ad ogni istante può generare un simbolo tra quelli compresi tra i caratteri alfanumerici.

In base a tale distinzione si ha la classificazione seguente:

- sistemi di comunicazione analogici: sistemi che trasferiscono informazione tra una sorgente analogica e l'utente;
- sistemi di comunicazione discreti o digitali: sistemi che trasferiscono informazione tra una sorgente discreta e l'utente;
- sistemi di comunicazione ibridi: sistemi che trasferiscono informazione tra una sorgente analogica e l'utente utilizzando tecniche proprie dei sistemi di comunicazione digitali.

Usualmente la sorgente e l'utente sono in punti diversi dello spazio, spesso molto distanti tra loro. Il canale di trasmissione è quell'oggetto che fornisce la connessione tra di essi. Il canale può essere di differente natura. In particolare possiamo distinguere:

- canali hardwire: linee telefoniche, cavi coassiali, fibre ottiche, ecc.;
- canali software: atmosfera, spazio esterno, acqua, ecc.

Il canale è una parte del sistema di comunicazione che può non essere progettabile, ma imposta dalla particolare applicazione. Ad esempio, per il collegamento con un aereo, l'unica possibilità è l'utilizzo dell'atmosfera come canale di trasmissione. Il canale è generalmente sorgente di degradazione, attenuazione e interferenza per il segnale trasmesso, in modo dipendente dalle caratteristiche del canale stesso e del particolare segnale inviato attraverso di esso. Stabilito il canale di trasmissione, un sistema di comunicazione dovrà far transitare l'informazione attraverso di esso in modo da fargli subire la minore distorsione possibile. Questo è lo scopo del trasmettitore. Dalla parte dell'utente, il ricevitore si occuperà di determinare il

segnale informativo originario dal segnale in uscita dal canale nel modo più fedele possibile. Un particolare canale può richiedere che il segnale trasmesso abbia particolari caratteristiche per transitarvi con la minore distorsione possibile. In linea di principio se abbiamo a disposizione un certo canale e conosciamo un particolare segnale che può transitarvi, possiamo pensare di “sovrapporre” ad esso il segnale informativo. Questa operazione prende il nome di modulazione, ed è realizzata dal blocco modulatore di Fig. 4.1.

Tramite un processo duale, la demodulazione, il ricevitore estrae il segnale informativo a partire dal segnale ricevuto, degradato dagli effetti indesiderati introdotti dal canale (attenuazione, rumore, interferenza, ecc.). Tale degradazione può ripercuotersi sul segnale informativo estratto e inviato all’utente. Per un dato tipo di modulazione ci possono essere diverse tecniche di demodulazione, che comportano complessità, accuratezza nel ricostruire il segnale informativo inviato e costi diversi.

4.1.1 Sistemi di comunicazione digitale

Nel caso di un sistema di comunicazione digitale la sorgente informativa è di tipo discreto.

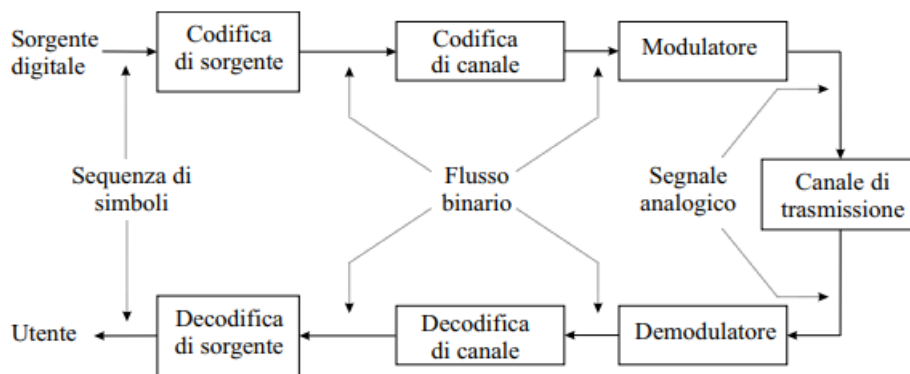


Figura 4.2: Blocchi fondamentali di un sistema di comunicazione digitale

I blocchi funzionali di un sistema di comunicazione digitale sono rappresentati in Fig. 4.2. In particolare notiamo la presenza di due nuovi blocchi: il codificatore di sorgente ed il codificatore di canale, a cui corrispondono i blocchi duali di decodifica di sorgente e di canale al ricevitore. Il codificatore di sorgente provvede a rappresentare in un alfabeto binario i simboli in uscita dalla sorgente informativa. Nota la statistica con cui si presentano tali simboli, è possibile progettare la codifica di sorgente per minimizzare il numero di bit al secondo da trasmettere. Al contrario, il codificatore di canale introduce ridondanza (bit aggiuntivi non informativi) in maniera controllata per consentire la rilevazione e la correzione di un certo numero di eventuali errori di trasmissione provocati dalla degradazione del segnale introdotta dal canale. Ciò avviene a spese di una complicazione del sistema e di una riduzione del flusso informativo, ci sono più bit da trasmettere e quindi meno bit informativi al secondo.

4.2 DTMF

Il DTMF è un sistema di segnalazione per identificare le chiavi o meglio dire il numero composto da un pulsante o da una tastiera DTMF. Esso è un sistema di chiamata a toni multifrequenza che ha abilitato la segnalazione a lunga distanza dei numeri compiuti nella gamma di frequenze vocali sulle linee telefoniche. Ciò ha eliminato la necessità di un operatore di telecomunicazioni tra il chiamante e il ricevitore e ha evoluto la selezione automatica nei centri di commutazione telefonica.

I segnali DTMF sono stati sviluppati dagli AT&T Bell Labs e sono una specie di segnali audio applicati alla rete telefonica. Il sistema DTMF è caratterizzato dall'uso dell'analisi nel dominio della frequenza.

I Bell Labs hanno selezionato due set di frequenze, quelle basse (697Hz, 770Hz, 852Hz, 941Hz) e quelle alte (1209Hz, 1336Hz, 1477Hz, 1633Hz); i DTMF (Dual Tone Multi Frequency) come suggerisce il nome, usano una combinazione di due toni d'onda sinusoidale per rappresentare una chiave. Questi toni sono chiamati frequenze di riga e colonna in quanto corrispondono al layout di una tastiera telefonica.

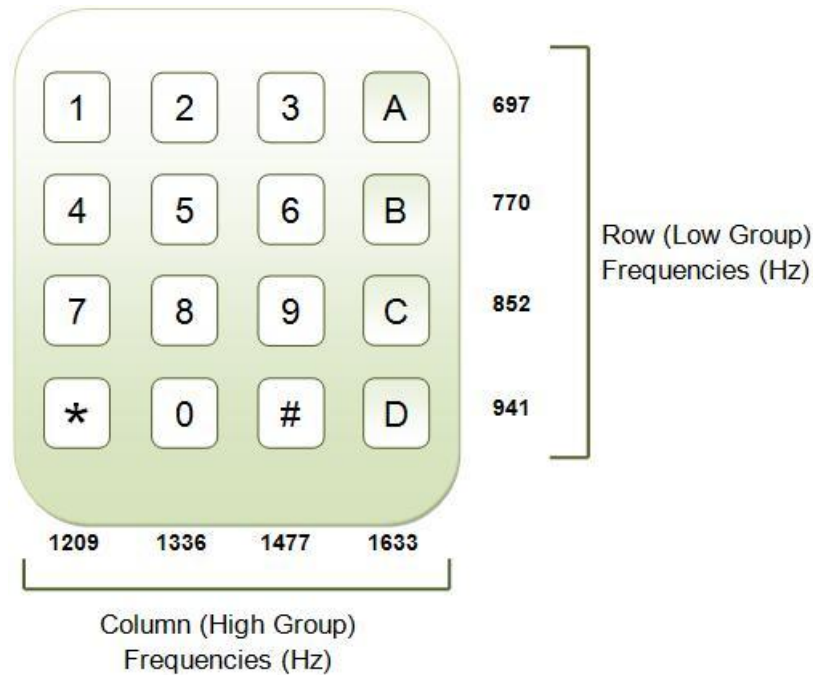


Figura 4.3: Keyboard layout DTMF

Una tastiera DTMF (generatore o encoder) genera un tono sinusoidale che è un mix di una frequenza di riga e una di colonna. Le frequenze di riga sono basse frequenze, le frequenze delle colonne appartengono alle frequenze elevate. Ciò impedisce l'errata interpretazione delle armoniche. Esse sono scelte in modo tale che nessuno ha una relazione armonica con gli altri e che la miscelazione delle frequenze non produrrà somme o frequenze di prodotto che potrebbero imitare un altro tono valido. Le frequenze di alto livello (i toni della colonna) sono leggermente più alti del gruppo basso per compensare il roll-out ad alta frequenza dei sistemi audio vocali. Le frequenze di riga e colonna corrispondenti a una tastiera DTMF sono state indicate nella figura precedente. I toni DTMF sono in grado di rappresentare uno

dei 16 stati o simboli diversi sulla tastiera. Questo equivale a 4 bit di dati, noto anche come nibble.

4.3 Audio Metadata

I metadati musicali, anche comunemente denominati metadati ID3, sono informazioni incorporate in un file audio, utilizzate per identificarne il contenuto. Questi dati che sono nella maggior parte dei file della libreria musicale digitale, possono essere utilizzati da una vasta gamma di dispositivi elettronici di consumo e programmi software. La ragione più comune per utilizzare metadati incorporati in un file audio digitale è per scopi di identificazione.

I dettagli di una canzone, ad esempio, possono essere visualizzati durante la riproduzione per facilitarne l'identificazione. A seconda del formato audio utilizzato, esiste un'area speciale (normalmente all'inizio o alla fine del file) riservata ai metadati che identifica l'audio codificato in diversi modi. Queste informazioni possono essere utili per la gestione e l'organizzazione della libreria. Esempi di tipi di informazioni che possono essere memorizzati nell'area metadati di un file audio includono:

- Il titolo della canzone.
- La band o il nome dell'artista.
- Il nome dell'album.
- Tipo di musica (genere).
- Numero di traccia dell'album.
- Anno in cui la canzone è stata rilasciata.

Per esempio, per il formato MP3, esistono due metadati utilizzati per la codifica dei file audio. Questi sono chiamati ID3v1 e ID3v2 - da questo deriva il termine ID3 tags. La prima versione di ID3 (v1) memorizza le informazioni alla fine di un file MP3 con spazio allocato per un massimo di 128 byte di dati.

D'altra parte, la versione 2 [27] (ID3v2) si trova all'inizio di un file MP3 ed è un formato contenitore basato su frame. È molto più capace e ha una capacità molto più grande per memorizzare metadati - fino a 256 MB.

Un altro formato di metadati è l'Extensible Metadata Platform (XMP), uno standard ISO, originariamente creato da Adobe Systems Inc., per la creazione, l'elaborazione e lo scambio di metadati standardizzati e personalizzati per documenti digitali e set di dati. XMP standardizza un modello di dati, un formato di serializzazione e proprietà di base per la definizione e l'elaborazione di metadati estensibili. Il modello di dati XMP, il formato di serializzazione e le proprietà principali sono pubblicate dall'Organizzazione internazionale per la standardizzazione come standard ISO 16684-1: 2012 [15].

Il modello di dati XMP può essere utilizzato per memorizzare qualsiasi insieme di proprietà dei metadati. Queste possono essere semplici coppie nome-valore, valori strutturati o liste di valori. I dati possono essere annidati. Lo standard XMP definisce anche spazi dei nomi specifici per set definiti di proprietà. Gli spazi dei nomi personalizzati possono essere utilizzati per estendere il modello dei dati.

Il modello di dati XMP astratto richiede una rappresentazione concreta quando viene memorizzato o incorporato in un file. Come formato di serializzazione, viene utilizzato un sottoinsieme della sintassi W3C RDF/XML. Essa è una sintassi per esprimere un grafico di struttura delle risorse in XML. Esistono diversi modi equivalenti per serializzare lo stesso pacchetto XMP in RDF/XML. I tag metadati più comuni registrati nei dati XMP sono quelli della Dublin Core Metadata Initiative, che comprendono cose come titolo, descrizione, autore e così via. Lo standard è stato progettato per essere estensibile, consentendo agli utenti di aggiungere i propri metadati personalizzati nei dati XMP. XMP in genere non consente di incorporare tipi di dati binari. Ciò significa che tutti i dati binari che si desidera trasportare in XMP, come le immagini in miniatura, devono essere codificati in un formato XML adatto, ad esempio Base64.

I metadati XMP possono descrivere un documento nel suo complesso, ma possono anche descrivere parti di un documento, ad esempio pagine o immagini incluse. Questa architettura consente di conservare informazioni sugli autori e sui diritti, ad esempio sulle immagini incluse in un documento pubblicato. Allo stesso modo, consente ai documenti creati da diversi documenti più piccoli di conservare i metadati originali associati alle parti.

4.4 Audio encoder

Il compito fondamentale di un sistema di codifica audio percettivo è quello di comprimere i dati audio digitali in modo che:

- la compressione sia la più efficiente possibile, cioè il file compresso deve essere il più piccolo possibile.
- i suoni ricostruiti (decodificati) suonino esattamente (o il più vicino possibile) all'audio originale.

Altri requisiti per la compressione audio includono la bassa complessità (per consentire ai decodificatori software o hardware un basso costo con un basso consumo energetico) e la flessibilità per diversi scenari di applicazione.

La tecnica per farlo è chiamata codifica percettiva e utilizza conoscenze dalla psicoacustica per raggiungere l'obiettivo di una compressione efficiente.

La codifica percettiva è una tecnica di compressione a perdita, cioè il file decodificato non è una replica esatta dei dati audio digitali originali. Ai fini di questo documento ci concentreremo sul formato utilizzato per Internet e dispositivi audio portatili basati su memoria audio e flash, MPEG-1/2 Layer-3 (aka MP3) e il formato che alla fine sarà il successore di Layer-3, vale a dire MPEG-2 Advanced Audio Coding (AAC). La figura 4.4 mostra lo schema a blocchi di base di un sistema di codifica percettiva costituito dai seguenti elementi:

- **Filter Bank:** Un filter bank viene utilizzato per decomporre il segnale di ingresso in componenti spettrali susseguenti (nel dominio del tempo e della frequenza). Insieme al corrispondente filter bank nel decodificatore forma un sistema analisi/sintesi.
- **Perceptual model:** Usando il segnale di ingresso nel dominio del tempo e/o l'output del filter bank di analisi, viene calcolata una stima della soglia effettiva

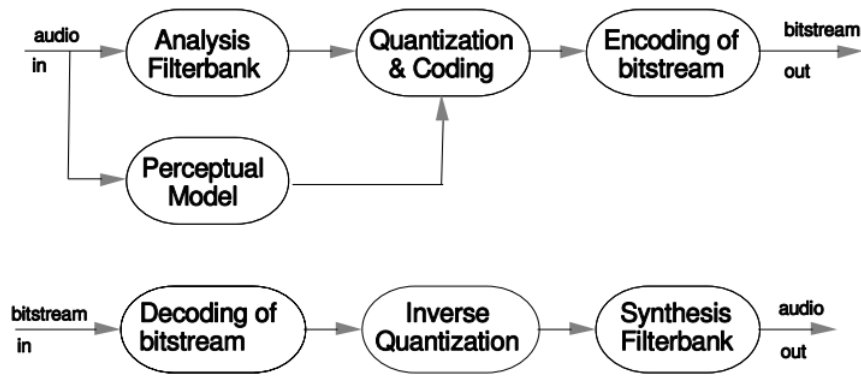


Figura 4.4: Diagramma a blocchi di un sistema di codifica/decodifica percettiva.

di mascheramento (tempo e frequenza) usando regole note dalla psicoacustica. Si definisce così il modello percettivo del sistema.

- **Quantization & Coding:** Le componenti spettrali sono quantizzate e codificate con l'obiettivo di mantenere il rumore, che viene introdotto dalla quantizzazione al di sotto della soglia di mascheramento. A seconda dell'algoritmo, questo passo avviene in modi molto diversi, dal semplice blocco a sistemi di analisi per-sintesi che utilizzano compressioni senza rumori aggiuntive.
- **Encoding of bitstream:** Un bitstream formatter viene utilizzato per assemblare lo stream di bit, tipicamente costituito dai coefficienti spettrali quantizzati e codificati e da alcune informazioni laterali, ad es. informazioni sull'allocazione dei bit.

La seguente descrizione della codifica Layer-3 mette a fuoco le funzioni di base e un certo numero di dettagli necessari per comprendere le implicazioni delle opzioni di codifica sulla qualità del suono. Non è una descrizione completa di come creare un encoder MPEG-1 Layer-3. Per essere applicabile a diversi scenari di applicazione, MPEG ha definito una rappresentazione dei dati che comprende una serie di opzioni.

- **Operating mode:** L'audio MPEG-1 funziona sia per segnali mono che stereo. Una tecnica chiamata codifica joint stereo può essere utilizzata per rendere più efficiente la codifica combinata dei canali sinistro e destro di un segnale audio stereofonico. Layer-3 consente sia la codifica stereo mid/side e, per i bit-rate più bassi, la codifica stereo di intensità. La codifica stereo di intensità consente di ridurre i bitrate, ma porta il rischio di cambiare l'immagine audio. I modi operativi sono:
 - Single Channel
 - Dual Channel (due canali indipendenti, per esempio contenenti diverse versioni linguistiche dell'album)
 - Stereo
 - Joint Stereo
- **Sampling Frequency:** La compressione audio MPEG funziona su una serie di frequenze di campionamento diverse. MPEG-1 definisce la compressione audio

a 32 kHz, 44,1 kHz e 48 kHz. MPEG-2 estende questo a metà delle frequenze, cioè 16 kHz, 22.05 e 24 kHz. MPEG-2.5 è il nome di una estensione proprietaria di Fraunhofer e MPEG-1/2 Layer-3 e funziona a frequenze di campionamento di 8 kHz, 11.05 e 12 kHz.

- Bit-rate: L'audio MPEG non funziona solo a un rapporto di compressione fisso. La selezione del bit-rate dell'alimentazione compressa è, entro certi limiti, completamente lasciata all'attuatore o operatore di un codificatore audio MPEG. Lo standard definisce una gamma di bitrate da 32 kbit/s (nel caso di MPEG-1) o 8 kbit/s (nel caso dell'estensione MPEG-2 Low Sampling Frequencies (LSF) fino a 320 kbit/s risp. 160 kbit/s per LSF). Nel caso di MPEG-1/2 Layer-3, la commutazione di bit rate da frame a frame deve essere supportata da decodificatori.

4.4.1 Descrizione dell'algoritmo

I paragrafi seguenti descrivono l'algoritmo di codifica Layer-3 lungo i blocchi di base di un encoder percettivo. Maggiori informazioni su Layer-3 possono essere trovate in [10] e [11]. La figura 5.15 mostra lo schema a blocchi di un tipico encoder MPEG-1/2 Layer-3.

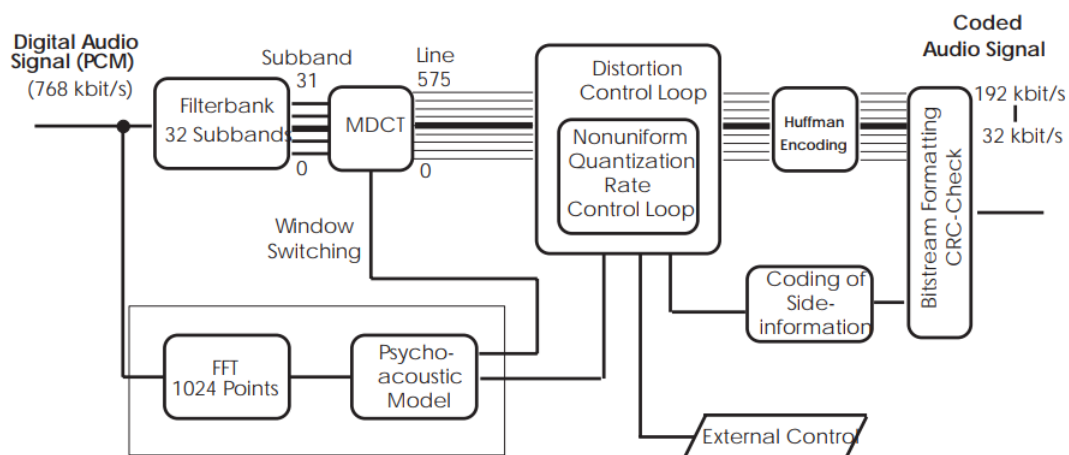


Figura 4.5: Diagramma a blocchi di un MPEG-1 Layer-3 encoder.

La filterbank utilizzata in MPEG-1 Layer-3 appartiene alla classe delle filter bank ibride. Questo filtro viene costruito eseguendo la cascata di due tipi diversi di filter bank: il filtro polyphase (come utilizzato in Layer-1 e Layer-2) e successivamente il Modified Discrete Cosine Transform (MDCT) filter. Il filtro polyphase ha lo scopo di rendere Layer-3 più simile a Layer-1 e Layer-2. La suddivisione di ciascuna banda di frequenza polifase in 18 subfidi più sottili aumenta il potenziale di rimozione della ridondanza, portando ad una migliore efficienza di codifica dei segnali tonali. Un altro risultato positivo di una migliore risoluzione di frequenza è il fatto che il segnale di errore può essere controllato per consentire un monitoraggio più fine della soglia di mascheramento. Il filter bank può essere commutato a una risoluzione di frequenza minore per evitare le preechoes (vedi sotto).

Il modello percettivo utilizza entrambi i filter bank come descritto in [23] e unisce il calcolo dei valori di energia (per i calcoli di mascheramento) e la filter bank prin-

cipale. L'output del modello percettivo consiste nei valori per la soglia di mascheramento o dal rumore consentito per ogni partizione del codificatore. Nel Layer-3, queste partizioni del codificatore sono approssimativamente equivalente alle fasce critiche dell'udito umano. Se il rumore di quantizzazione può essere mantenuto sotto la soglia di mascheramento per ogni partizione del codificatore, il risultato di compressione deve essere indistinguibile dal segnale originale.

Un sistema di due cicli di iterazione nidificati è la soluzione comune per la quantizzazione e la codifica in un encoder Layer-3. La quantizzazione avviene tramite un quantizzatore power-law. In questo modo, i valori più grandi vengono codificati automaticamente con una minore accuratezza e una certa forma di rumore è già incorporata nel processo di quantizzazione. I valori quantizzati sono codificati dalla codifica Huffman. Per adattare il processo di codifica a diverse statistiche locali dei segnali musicali, la tabella ottimale di Huffman è selezionata da una serie di scelte. La codifica Huffman funziona su coppie e, solo nel caso di numeri molto piccoli da codificare, quadruplica. Per ottenere un adattamento ancora migliore alle statistiche dei segnali, diverse tabelle di codice Huffman possono essere selezionate per diverse parti dello spettro. Il processo per trovare il guadagno e gli scalatori ottimali per un dato blocco, il bit-rate e l'output del modello percettivo è di solito fatto da due cicli di iterazione nidificati in un modo di analisi per-sintesi:

- Inner iteration loop (rate loop): Le tabelle di codice Huffman assegnano parole di codice più brevi a valori (più frequenti) minimi quantizzati. Se il numero di bit risultante dall'operazione di codifica supera il numero di bit disponibili per codificare un dato blocco di dati, questo può essere corretto regolando il guadagno globale per ottenere una maggiore grandezza di quantizzazione, portando a valori minimi quantizzati. Questa operazione viene ripetuta con diverse grandezze di quantizzazione finché la richiesta di bit risultante per la codifica Huffman non è sufficiente. Il loop viene chiamato loop di velocità poiché modifica il tasso di codifica complessivo finché non è abbastanza piccolo.
- Outer iteration loop (noise control loop): Per formare il rumore di quantizzazione in base alla soglia di mascheramento, i fattori di scala vengono applicati a ciascuna fascia di fattore di scala. I sistemi iniziano con un fattore predefinito di 1,0 per ciascuna banda. Se il rumore di quantizzazione in una data banda si trova a superare la soglia di mascheramento (rumore consentito) fornito dal modello percettivo, il fattore di scala di questa banda viene regolato per ridurre il rumore di quantizzazione. Dato che il raggiungimento di un rumore di quantizzazione minore richiede un maggior numero di passi di quantizzazione e quindi un bit-rate più elevato, il ciclo di regolazione della velocità deve essere ripetuto ogni volta che vengono utilizzati nuovi fattori di scala. In altre parole, il ciclo di frequenza è annidato all'interno del circuito di controllo del rumore. Il loop esterno (rumore) viene eseguito finché il rumore reale (calcolato dalla differenza dei valori spettrali originali meno i valori spettrali quantizzati) è inferiore alla soglia di mascheratura per ogni banda del fattore di scala (cioè la banda critica).

Mentre il ciclo di iterazione interno converge sempre (se necessario, impostando la dimensione di fase di quantizzazione abbastanza grande per eliminare tutti i valori spettrali), ciò non è vero per la combinazione di entrambi i cicli di iterazione. Se il modello percettivo richiede dimensioni di quantizzazione così piccole che

il loop di velocità deve sempre aumentarle per abilitare la codifica alla velocità di bit richiesta, entrambi possono continuare per sempre. Per evitare questa situazione, possono essere verificate diverse condizioni per arrestare anticipatamente le iterazioni. Tuttavia, per la codifica veloce e per i buoni risultati di codifica, questa condizione dovrebbe essere evitata. Questo è un motivo per cui un encoder MPEG Layer-3 (lo stesso vale per AAC) ha generalmente bisogno di sintonizzare i set di parametri del modello percettivo per ogni bit rate.

La figura 4.6 mostra uno schema a blocchi di un encoder MPEG-2 AAC. AAC segue lo stesso paradigma di codifica di base come Layer-3 (filtro a risoluzione ad alta frequenza, quantizzazione non uniforme, codifica Huffman, struttura di loop di iterazione utilizzando analisi per sintesi), ma migliora su Layer-3 in molti dettagli e utilizza nuovi strumenti di codifica per migliorare la qualità a bassi bit rate.

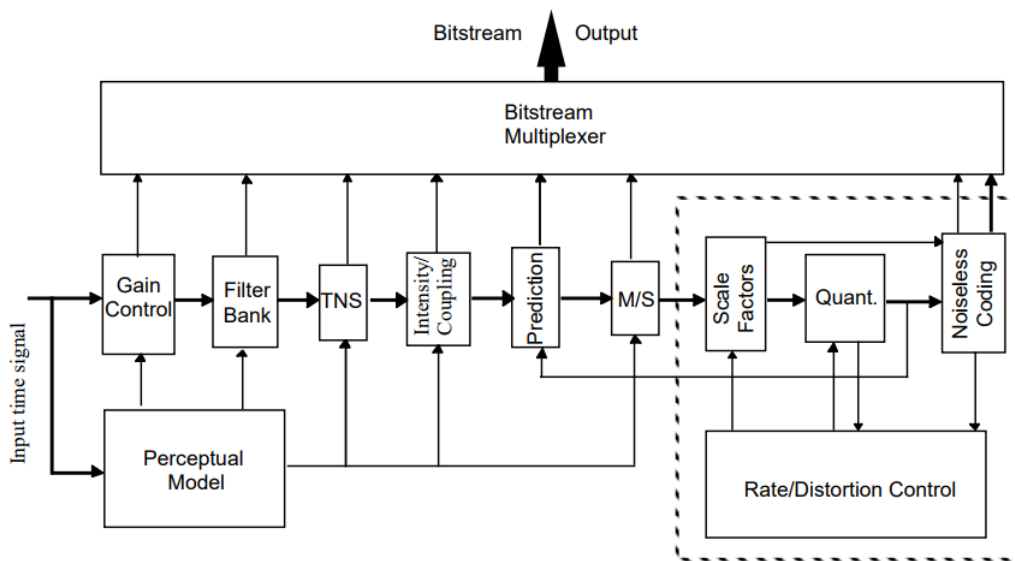


Figura 4.6: Diagramma a blocchi di un MPEG-2 AAC encoder.

4.4.2 Strumenti per migliorare l'efficienza di codifica

Le seguenti modifiche rispetto a Layer-3 aiutano a ottenere la stessa qualità a bitrate più bassi:

- Risoluzione di frequenza superiore : Il numero di linee di frequenza in AAC è fino a 1024 rispetto a 576 per Layer-3
- Migliore codifica joint stereo: Rispetto a Layer-3, sia la codifica mid/side che la codifica dell'intensità sono più flessibili, consentendo di ridurre più frequentemente il bit rate.
- Migliore codifica di Huffman: n AAC, la codifica per quadrupli di linee di frequenza viene applicata più spesso. Inoltre, l'assegnazione di tabelle di codice Huffman alle partizioni del codificatore può essere molto più flessibile.

La codifica audio MPEG non funziona con una velocità di compressione fissa. L'utente può scegliere il bit-rate e così il fattore di compressione. Bassi livelli di bit-rate porteranno a fattori di compressione più elevati, ma qualità inferiore dell'audio

compresso. Bit-rate più elevati portano ad una minore probabilità di segnali con qualsiasi artefatto udibile. Tuttavia, diversi algoritmi di codifica hanno "sweet spot" dove funzionano meglio. A bit-rate molto più grandi di questo bit-rate target la qualità audio migliora molto lentamente e, a bit-rate molto più bassi la qualità diminuisce molto velocemente. Il "sweet spot" dipende dalle caratteristiche del codec come i codici Huffman. Per il Layer-3 questo bit di destinazione è di circa 1,33 bit/campione (vale a dire 128 kbit/s per un segnale stereo a 48 kHz), per AAC è circa 1 bit/campione (ossia 96 kbit/s per un segnale stereo a 48 kHz). A causa della codifica Huffman più flessibile, AAC può mantenere l'efficienza di codifica di base fino a tassi di bit più elevati che consentono qualità superiori. La codifica multicanale, grazie alle tecniche di codifica stereo articolate impiegate, è un pò più efficiente per campione rispetto a quella stereo e alla codifica mono. Si può trovare una buona panoramica del compromesso tra i bit-rate e la qualità ottenibile per un certo numero di algoritmi di codifica (inclusi AAC e MP3) [1].

È certamente vero che un gran numero di soggetti (soprattutto giovani) sono perfettamente in grado di ascoltare suoni singoli a frequenze ben al di sopra di 20 kHz. Tuttavia, contrariamente alla credenza popolare, non si è a conoscenza di alcun esperimento scientifico che ha dimostrato senza dubbio l'esistenza di un ascoltatore (addestrato o no) in grado di rilevare la differenza tra un segnale musicale (complesso) con contenuto fino a 20 kHz e lo stesso segnale, ma limitato a circa 16 kHz. Per chiarire, ci sono alcuni suggerimenti sul fatto che ci sono ascoltatori con tali capacità, ma la prova scientifica completa non è ancora stata data. Come corollario di questo teorema, è una buona strategia di codifica, limitare la risposta in frequenza di un encoder MP3 o AAC a 16 kHz (o più in basso se necessario).

Mentre la perdita di larghezza di banda sotto la frequenza data dai limiti dell'udito umano è un artefatto di codifica, non è necessariamente vero che un encoder che produce un audio compresso con maggiore ampiezza di banda suoni meglio. C'è un compromesso fondamentale dove passare i bit disponibili per la codifica. Se vengono utilizzati per migliorare la risposta in frequenza, non sono più disponibili per produrre un suono pulito a frequenze inferiori. Secondo lo stato attuale dell'arte, è meglio introdurre una limitazione di larghezza di banda fissa se la codifica viene fatta a un bit rate in cui non è possibile riprodurre in modo uniforme un segnale pulito con la larghezza di banda completa. Tecnicamente, sia MP3 che AAC possono riprodurre il contenuto del segnale fino al limite dato dalla frequenza reale di campionamento. Se ci sono encoder con una risposta fissa limitata di frequenza (ad una determinata bit rate) rispetto ad un altro encoder con larghezza di banda molto più ampia (allo stesso bit rate), l'esperienza dice che nella maggior parte dei casi l'encoder con la larghezza di banda più bassa produce un migliore audio compresso. Tuttavia, esiste un limite a questa affermazione: a bassi bit rate (64 kbit/s) la questione del miglior compromesso in termini di larghezza di banda rispetto alla pulizia è una questione fortemente contestata del gusto. Abbiamo scoperto che persino gli ascoltatori addestrati talvolta sono completamente in disaccordo circa la larghezza di banda che dovrebbe essere usata in un determinato encoder.

Utilizzando un encoder con buone prestazioni, sia Layer-3 che MPEG-2 Advanced Audio Coding possono comprimere la musica pur mantenendo la qualità vicino o uguale a quella del CD.

Tra i due codec, Layer-3 è il sistema di scelta per le applicazioni con qualità vicino

al CD. AAC è il suo successore designato, fornendo una qualità quasi CD a velocità di compressione più elevate (aumentando il tempo di riproduzione di dispositivi basati su memoria flash di quasi il 50% pur mantenendo la stessa qualità rispetto a Layer-3) e consentendo una codifica e una riproduzione di alta qualità fino ad alta definizione audio (a 96 kHz di frequenza di campionamento). AAC sarà il sistema di compressione scelto per la futura distribuzione elettronica di musica (EMD) che seguirà le tracce di altri algoritmi di compressione MPEG definiti in tutto il mondo come MPEG Audio Layer-2, MPEG Audio Layer-3 o MPEG Video[21].

Capitolo 5

Sviluppo del lavoro

I capitoli precedenti, hanno illustrato le motivazioni per l'utilizzo di un sistema di "Ad replacement". L'integrazione di questo tipo di sistema nelle web-radio è necessaria non solo per finalità economiche, come indicato nel capitolo 2, ma anche dal lato dell'ascoltatore. Una pubblicità mirata contribuisce ad aumentare l'user-experience.

Nei capitoli 3 e 4 sono stati esposti alcune tecniche e tecnologie audio che verranno utilizzate nelle soluzioni proposte in questo capitolo.

Vediamo ora le due implementazioni, la prima che fa uso dei metadati e la seconda che utilizza gli ultrasuoni.

5.1 Implementazione tramite metadati

In questa sezione viene esposta la soluzione che fa uso dei metadati al fine di rimpiazzare le pubblicità nel flusso audio di una web-radio.

Per iniziare verrà riproposta l'architettura radio vista nel capitolo 2 al fine di analizzarne i punti critici per poi proporre la soluzione adottata al fine di evitarli.

5.1.1 Architettura del sistema

Riprendendo la figura 5.1, illustriamo adesso i punti d'interesse al fine di utilizzare i metadati per la segnalazione.

Per prima cosa, andiamo ad analizzare l'elemento a monte della catena, cioè il software che si occupa di gestire la regia radio. Come già detto precedentemente, le soluzioni proposte, vogliono essere il meno invasive possibile. A questo scopo, il software di regia utilizzato nella soluzione proposta, è uno dei più comuni utilizzati dalle radio italiane.

Il programma in questione, raffigurato in figura 5.1, è Mb Studio [30], un potente software di automazione radiofonica. Il player audio è molto veloce, dispone di un controllo automatico del livello audio e di un compressore e supporta tutti i più comuni formati audio lineari e/o compressi (WAV, MP2, MP3, MP4, AAC, OGG, WMA, FLAC).

Il software è dotato di una programmazione commerciale che gestisce automaticamente tutte le scadenze contrattuali (data di inizio, fine, scelta dei giorni della settimana, programmazione a giorni e/o a settimane alterne, esclusione dai festivi, etc) e consente di impostare limiti all'affollamento pubblicitario orario e/o giornaliero;



Figura 5.1: Software di regia Mb Studio

lo stesso tipo di programmazione è applicabile anche ad appuntamenti informativi, rubriche e/o programmi. Pieno supporto anche per i contributi esterni provenienti da satellite: è possibile collegare MB STUDIO ai ricevitori satellitari professionali forniti da numerose agenzie o sindacations, sincronizzando la messa in onda in diretta oppure registrando i contenuti su disco per la ritrasmissione in differita. MB STUDIO può funzionare in un singolo computer desktop, in una rete di computers, o in configurazione client/server; non è richiesto hardware dedicato e sono supportate sia le schede audio consumer che quelle professionali.

Al fine di implementare questa soluzione, è necessario inserire nel flusso due tipi di metadata, il metadata "Spot" che indica l'immediato inizio di una pubblicità o di un carosello¹ e il metadata "Onair" che indica che il flusso audio seguente non è una pubblicità ma una normale diretta audio.

Il software precedentemente descritto, però, non permette l'inserimento arbitrario dei metadata appena citati quindi per risolvere questo problema è stato necessario inserire i metadata non direttamente nel software di regia, ma in un altro punto della catena cioè nell'encoder all'atto della codifica.

Per questo fine è stato necessario modificare l'architettura vista in figura 5.1, infatti, visto che non è possibile iniettare i metadata direttamente dal software, è necessario che il software di regia segnali all'encoder quando e quale metadata inserire.

¹carosello: sequenza di più pubblicità.

L'architettura viene modificata come illustrato in figura 5.2.

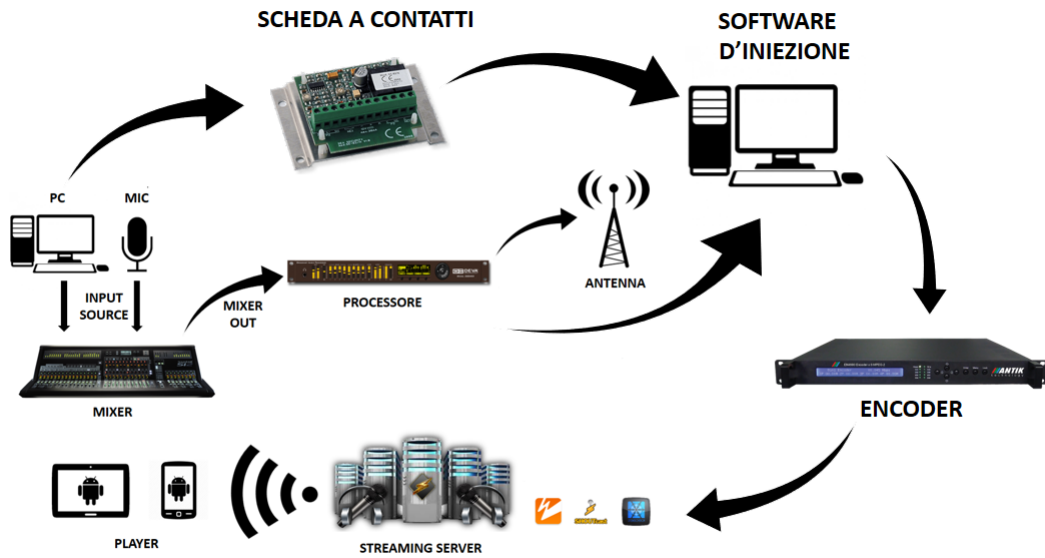


Figura 5.2: Architettura modificata per l'iniezione di metadata

Il software di regia permette di impostare dei bit di segnalazione su ogni brano, programma o pubblicità del palinsesto, all'inizio o alla fine dell'oggetto. La figura 5.3 mostra la finestra di impostazione, si può notare come sia possibile impostare diversi bit e quindi differenti segnali. Al nostro scopo verranno utilizzati solo due bit, uno per segnalare l'inizio di una pubblicità e uno per segnalarne la fine.



Figura 5.3: Finestra d'impostazione bit MbStudio

La figura 5.2 mostra un nuovo elemento, una scheda a contatti. Questa viene azionata dai bit generati dal software di regia, comunicati tramite una connessione TCP/IP ad un software realizzato di proposito il quale provvede ad inserire nel

preciso istante in cui il contatto della scheda si chiude, il relativo metadata nel flusso di codifica dell'encoder.

Come la figura 5.2 mostra, il software d'iniezione riceve, da un lato il flusso audio prodotto dal mixer ed elaborato dal processore, e dall'altro il segnale della scheda a contatti.

5.1.2 Funzionamento

Abbiamo visto l'introduzione di due nuovi attori, il software d'iniezione con annesso encoder, per l'inserimento dei metadata nel flusso streaming audio partendo dalle informazioni ricevute tramite l'altro attore di questa architettura, la scheda a contatti. Descriviamo ora il funzionamento dell'architettura modificata appena proposta. Per prima cosa, bisogna associare 2 bit ai metadata "Spot" e "Onair", in questo esempio useremo il bit 1 e il bit 2 della figura 5.3 per segnalare, rispettivamente, l'inserimento del metadata "Spot" e del metadata "Onair".

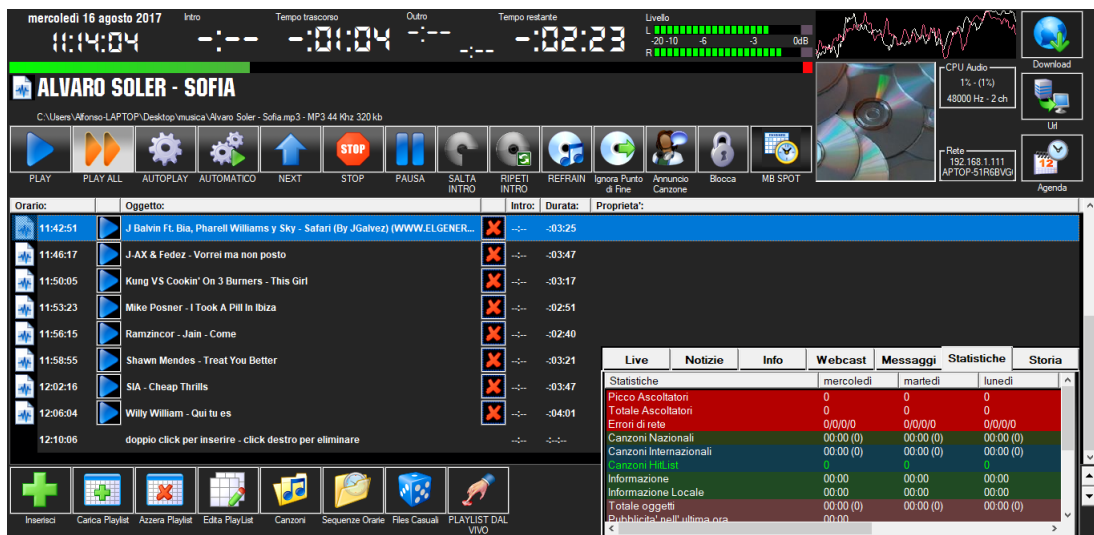


Figura 5.4: Esempio di palinsesto nel software MbStudio

In figura 5.4 è possibile vedere un esempio di palinsesto radio nel software di regia MbStudio. Il regista, a questo punto, affinché il sistema possa funzionare, deve impostare i due bit sopracitati all'inizio e alla fine di una pubblicità o di un carosello² pubblicitario. Per esempio, prendiamo come illustrazione di riferimento la figura 5.5, tra il brano denominato "Shawn Mendes -Treat You Better" e il brano denominato "SIA - Cheap Thrills" c'è il file contenente la pubblicità chiamato "pub mixdown". Su quest'ultimo il regista deve impostare i due bit, all'inizio della traccia attivare il bit 1 e alla fine attivare il bit 2.

²carosello: sequenza di più pubblicità.

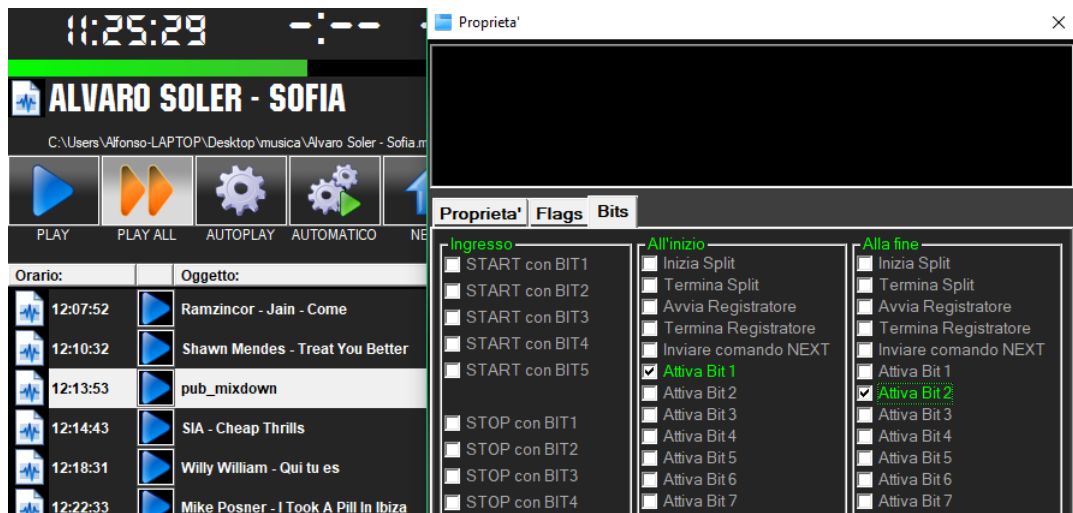


Figura 5.5: Esempio d'impostazione bit nel software MbStudio

Con questa impostazione, in fase d'esecuzione nel momento in cui andrà in onda la pubblicità, il software di regia attiverà il bit 1 che farà chiudere il contatto relativo della scheda, questo segnalerà al software d'iniezione l'inserimento del metadata "Spot" all'interno del file codificato.

Analogamente, alla fine della pubblicità, verrà attivato il bit 2 che chiuderà il contatto relativo sulla scheda e verrà inserito il metadata "Onair" che segna la fine di uno spazio pubblicitario.

Il software d'iniezione riceve il flusso audio dal processore e alla chiusura del contatto il segnale, proveniente dalla scheda, d'iniezione del metadata; il segnale d'iniezione, però, non è perfettamente sincronizzato con l'esatto punto in cui iniettare il metadata ma essendo questo ritardo costante è stato possibile aggiustarlo all'interno del software d'iniezione.

Infine, il flusso codificato, in uno dei formati di codifica più utilizzati come MP3 o ACC+, viene inviato al server di streaming che si occupa della distribuzione. In questa soluzione è stato utilizzato Wowza Streaming Engine (conosciuto anche come Wowza Media Server prima della versione 4) il quale è un software di media streaming unificato sviluppato da Wowza Media Systems. Il server è utilizzato per lo streaming di video, audio e applicazioni Internet su reti IP, su computer desktop, laptop e tablet, dispositivi mobili e altri dispositivi connessi in rete. Il server è un'applicazione Java distribuibile nella maggior parte dei sistemi operativi.[32]

Wowza è in grado di trasmettere contemporaneamente su più client e dispositivi di riproduzione, tra cui Adobe Flash Player, Microsoft Silverlight, Apple QuickTime Player e dispositivi iOS (iPad, iPhone, iPod Touch) e telefoni cellulari 3GPP.

Il motore Wowza Streaming è compatibile con i protocolli standard di streaming. Questi includono RTMP (e le varianti RTMPS, RTMPT, RTMPE, RTMPTE), HDS, HLS, MPEG DASH, RTSP, Smooth Streaming e MPEG-TS (unicast e multicast). Inoltre supporta anche flussi in entrata tramite protocolli RTSP e WOWZ provenienti da dispositivi mobili Android e iOS che eseguono l'applicazione di codifica mobile Wowza GoCoder.

Per lo streaming on-demand, Wowza Streaming Engine può gestire più tipi di file audio e video. I tipi di file supportati includono MP4 (contenitore QuickTime - .mp4, .f4v, .mov, .m4a, .m4v, .mp4a, .mp4v, .3gp e .3g2), FLV (Flash Video - .flv)

e contenuto MP3 (.mp3).

Il suo utilizzo, è stato scelto anche grazie a questa sua ultima caratteristica, infatti, grazie alla possibilità di fare streaming on-demand, l'applicazione descritta nel paragrafo seguente, è in grado di riprodurre il flusso anche in modo "time-shifted", spostato nel tempo e non solamente in diretta.

Ricapitolando, il software di regia, MbStudio in questo esempio, genera un segnale all'inizio di una inserzione pubblicitaria; questo segnale passa da un canale secondario, rispetto a quello del flusso audio che attraversa il processore. Il segnale chiude un contatto su una scheda che pilota il software d'iniezione. Quando il segnale si attiva, il software che di default esegue solo la codifica audio, genera un metadata di tipo "Spot" e lo inietta nella codifica. Il flusso codificato raggiunge lo streaming server che lo distribuisce in rete.

Analogamente, quando l'inserzione finisce, il software di regia genera un secondo segnale, diverso dal primo, che fa scattare un altro contatto sulla scheda che piloterà il software di codifica in modo da generare ed iniettare nel flusso il metadata "Onair".

5.1.3 Applicazione android

Vista la catena architetturale modificata al fine di inserire adeguatamente i metadata che verranno utilizzati dall'applicazione per riconoscere uno spot pubblicitario, vediamo ora come è strutturata l'applicazione, per dispositivi Android, in grado di effettuare la riproduzione del flusso audio e di effettuare la sostituzione della pubblicità quando segnalato dai metadata.

Il prototipo dell'applicazione è composto essenzialmente da due sezioni, la scelta del profilo e il player audio.

Uno dei passi preliminari è quello della scelta del profilo, in questo prototipo, per semplicità, è stato scelto come metodo di profilazione dell'utente il metodo esplicito (vedi par. 2.8.1).



Figura 5.6: User interface per la selezione del profilo

La figura 5.6 mostra l'interfaccia con cui l'utente interagisce per la selezione del profilo; in questo prototipo, è possibile scegliere tre gruppi di appartenenza differenti: uomo, donna e ragazzi.

Tramite questa scelta l'applicazione è in grado di utilizzare il gruppo d'appartenenza adatto all'utente. Selezionando uno dei tre target di appartenenza, l'applicazione tramite una richiesta http, scarica una lista di spot pubblicitari adatti al profilo selezionato e la memorizza sulla memoria locale del dispositivo in cui è in esecuzione l'applicazione.

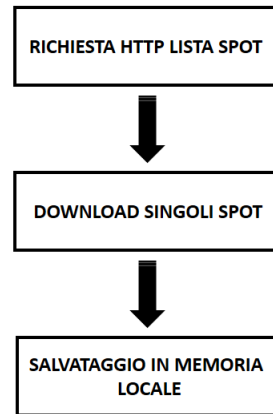


Figura 5.7: Flow chart download spot pubblicitari

La figura 5.7 illustra la sequenza di operazioni per il download degli spot pubblicitari.

Una volta configurato il gruppo di appartenenza e scaricati gli spot pubblicitari che andranno a sostituire quelli nel normale flusso di streaming, l'applicazione è pronta a riprodurre il flusso audio. A questo fine è stato scelto ExoPlayer, un lettore multimediale a livello applicazione per Android. Esso fornisce un'alternativa all'API MediaPlayer di Android per la riproduzione di audio e video sia a livello locale che su Internet. ExoPlayer supporta funzioni non attualmente supportate dall'API MediaPlayer di Android, incluse le riproduzioni adattive DASH³ e SmoothStreaming⁴. A differenza dell'API di MediaPlayer, ExoPlayer è facile da personalizzare e estendere e può essere aggiornato tramite aggiornamenti di applicazioni Play Store. [18]. ExoPlayer offre numerosi vantaggi rispetto al MediaPlayer integrato da Android. Oltre a supportare il DASH e lo SmoothStreaming, entrambi non supportati dal MediaPlayer di Android, supporta molti altri formati, supporta le caratteristiche avanzate di HLS⁵, ha la capacità di unire perfettamente, concatenare e di ciclare i media e la capacità di aggiornare il lettore insieme all'applicazione. Poiché ExoPlayer è una libreria inclusa nell'applicazione apk, è possibile controllare la versione

³DASH: Dynamic Adaptive Streaming over HTTP, anche chiamato MPEG-DASH, è una tecnica adattativa di bitrate streaming che consente lo streaming di contenuti multimediali su Internet in alta qualità fornito da server Web HTTP tradizionali.

⁴SmoothStreaming: è un'estensione dei servizi multimediali IIS che consente lo streaming adattativo dei supporti in Silverlight e in altri client tramite HTTP. Lo SmoothStreaming offre un'esperienza di riproduzione di alta qualità che scala sulle reti di distribuzione dei contenuti.

⁵HLS: è un protocollo di comunicazione streaming di contenuti multimediali via HTTP implementato da Apple come parte dei suoi prodotti software QuickTime, Safari, macOS e iOS.

che si utilizza e potere facilmente aggiornare a una versione più recente come parte di un normale aggiornamento dell'applicazione. Supporta inoltre la crittografia comune di Widevine su Android 4.4 (livello API 19) e superiore, la possibilità di personalizzare ed estendere il lettore in base al proprio caso d'uso, infatti ExoPlayer è stato progettato specificamente in questo modo e consente di sostituire molti componenti con implementazioni personalizzate. Infine ha la capacità di integrare rapidamente una serie di librerie aggiuntive utilizzando le estensioni ufficiali. Ad esempio, l'estensione IMA rende facile monetizzare i propri contenuti utilizzando l'Interactive Media Ads SDK[19].

È importante notare che ci sono anche alcuni svantaggi, i componenti audio e video standard di ExoPlayer si basano sull'API MediaCodec di Android, che è stato rilasciato in Android 4.1 (livello API 16), quindi non funzionano sulle versioni precedenti di Android.

Al centro della libreria ExoPlayer c'è l'interfaccia ExoPlayer. Un ExoPlayer espone le funzionalità tradizionali del lettore multimediale ad alto livello quali la capacità di fare buffer dei media, il play, la pausa e la ricerca. Le implementazioni sono progettate per fare alcune ipotesi (e quindi imporre poche limitazioni) sul tipo di supporto che viene riprodotto, come e dove viene memorizzato e come viene eseguito. Invece di implementare direttamente il caricamento e la presentazione dei media, le implementazioni di ExoPlayer delegano questo lavoro a componenti iniettati quando un lettore viene creato o quando è pronto per la riproduzione.

L'applicazione presenta una semplice interfaccia con i comandi principali del player, il play, il pause e il live. Oltre al semplice live streaming l'applicazione permette di ascoltare la radio in differita.



Figura 5.8: User interface di riproduzione Exoplayer

In figura 5.8 è possibile vedere l'user interface, tramite il play button è possibile iniziare la riproduzione, il pause button permette di mettere in pausa lo streaming per poi riprenderlo dal momento in cui è stato interrotto, infine il live button per-

mette di ritornare al live streaming se si era in differita.

L'interfaccia mostra inoltre, a scopi dimostrativi, informazioni riguardanti lo stato del player, cioè se si sta riproducendo il programma principale oppure uno spot personalizzato o se si è in differita fornisce informazioni sul tempo di ritardo rispetto alla diretta. Infine è possibile visualizzare i classici metadata riguardanti l'autore e il titolo del brano che si sta riproducendo.

La figura 5.9 mostra i vari stati del player.

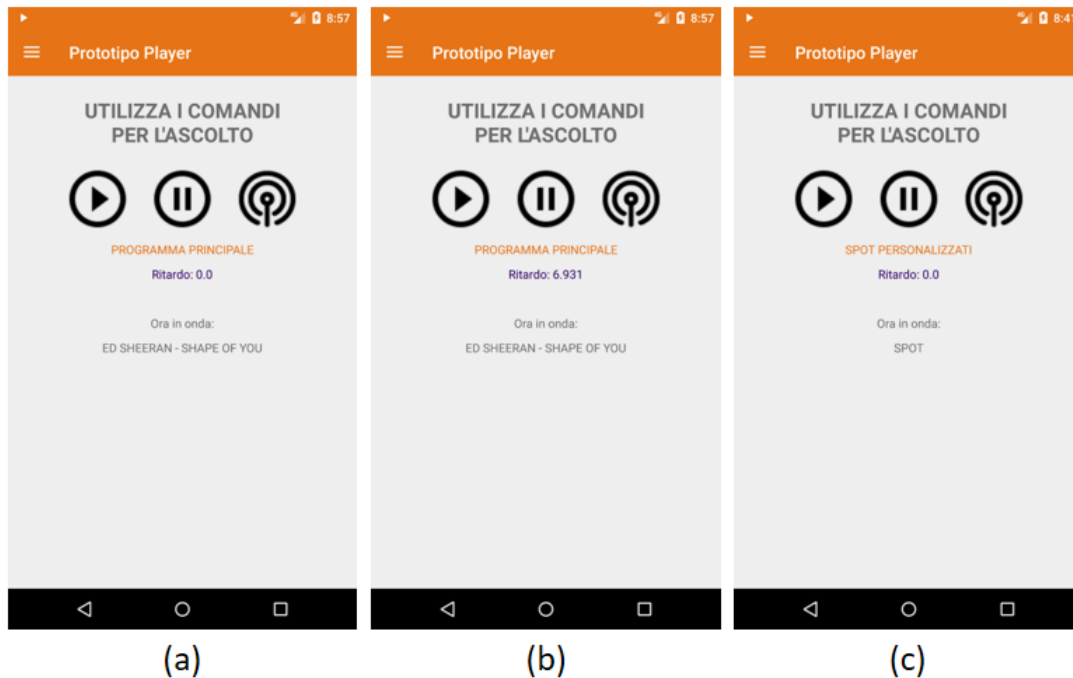


Figura 5.9: Stati del player (a) programma live (b) programma differita (c) spot personalizzato

La gestione del player è demandata ad un android Service [6] in esecuzione in background. Il service ospita due istanze di Exoplayer, una per riprodurre lo streaming live e uno in standby pronto a riprodurre gli spot locali personalizzati.

Il service di tipo bound consente ai vari componenti di un applicazione di essere legati insieme al fine di creare una connessione di lunga durata. In questo modo l'activity [3] che ospita la user interface è in grado di interagire con il service che a sua volta può esporre le sue funzionalità tramite la comunicazione inter-process (IPC). Il servizio espone i metodi per gestire il play, il pause e il live del player principale (quello che riproduce lo streaming radio) ed un'interfaccia per notificare all'activity un cambio di stato del player, inoltre, implementa l'interfaccia MetadataReader che notifica il servizio sulla presenza di un nuovo metadata. In questo metodo prende forma la logica del sistema, quando arriva un nuovo metadata, viene notificata l'activity sul nuovo metadata e in più se è presente una delle due keyword "Spot" o "Onair" viene attivato il sistema di scambio dei player. Il sistema di scambio prevede una dissolvenza incrociata (crossfade) tra i volumi dei due player, prendiamo in esempio l'arrivo del metadata "Spot", l'adv player viene immediatamente avviato mentre i volumi dei due player cominciano ad incrociarsi. Il volume dell'adv player sale in maniera esponenziale per evitare la perdita dei primi milli-

secondi dello spot mentre il volume dello streaming decresce logaritmicamente. La figura 5.10 mostra il grafico dei volumi durante il crossfade.

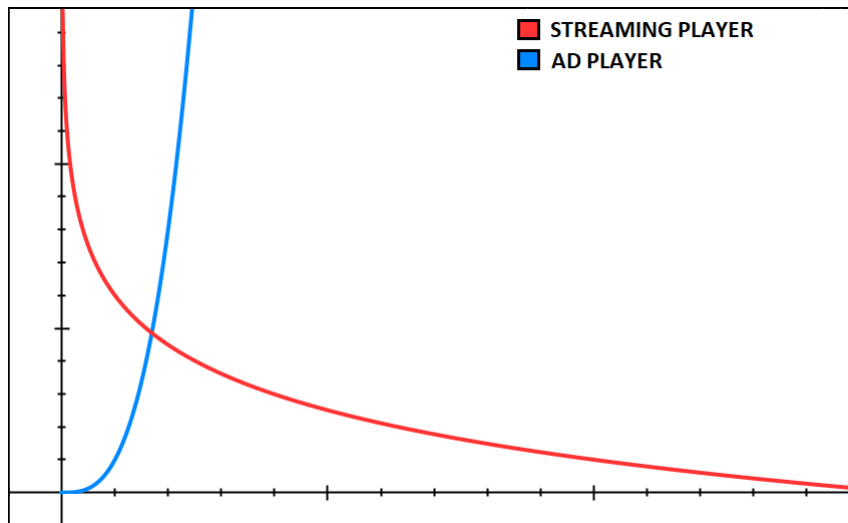


Figura 5.10: Grafico dei volumi durante il crossfade

Analogamente, alla ricezione del metadata "Onair", avviene un crossfade inverso dei volumi dei due player. L'applicazione permette inoltre, nella sua versione prototipale, lo spostamento del punto di crossfade, cioè l'incrocio delle due curve, nonché della durata complessiva del crossfade.

Infine, l'applicazione permette l'ascolto in differita quando si mette in pausa il player. Questo è possibile grazie alle caratteristiche del player e del server Wowza che supportano l'HLS (HTTP Live Streaming). HLS funziona come tutte le tecnologie di streaming adattativo; si creano più file per la distribuzione al lettore, che possono modificare in modo adattativo i flussi per ottimizzare l'esperienza di riproduzione. La distribuzione ai client HLS è possibile codificando l'origine in più file a velocità diverse, suddividendo il flusso in blocchi brevi (chunk), di solito tra 5 e 10 secondi. Questi vengono caricati sul server HTTP Wowza insieme a un file di manifesto basato su testo con un'estensione .M3U8 che indirizza il player a file di manifesto aggiuntivi per ciascuno dei flussi codificati. Come tecnologia basata su HTTP, tutta la logica di commutazione risiede sul lettore.

5.2 Implementazione tramite ultrasuoni

In questo paragrafo viene esposta la soluzione che fa uso degli ultrasuoni al fine di sostituire le pubblicità nel flusso streaming di una web-radio.

Come fatto per la precedente soluzione, verrà riproposta l'architettura radio vista nel capitolo 2 al fine di analizzarne i punti critici per poi proporre la soluzione adottata al fine di evitarli.

5.2.1 Architettura del sistema

La catena architetturale è essenzialmente quella già vista nella precedente soluzione e riproposta di seguito in figura 5.11. Il software di regia utilizzato per questa soluzione è sempre MBStudio, così come il resto dell'architettura che rimane invariata.

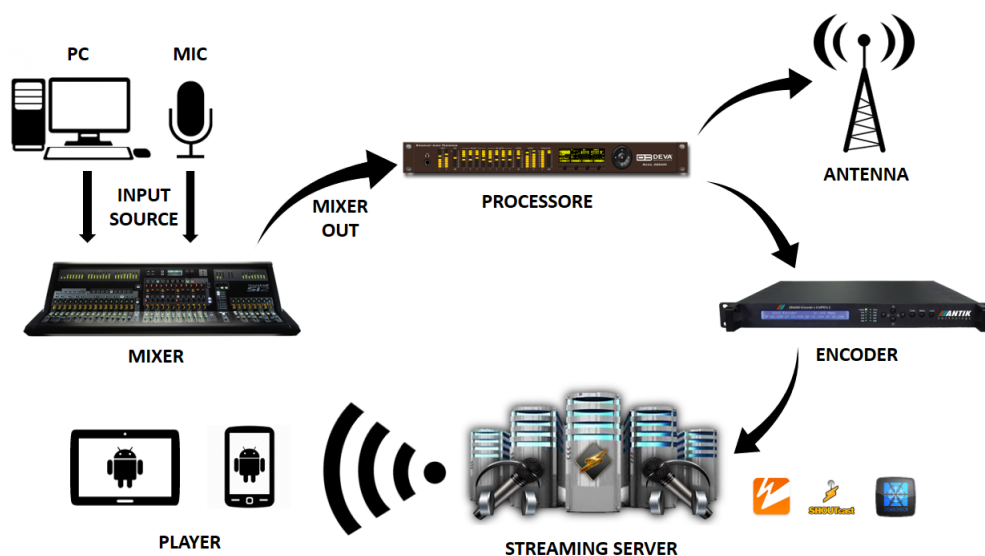


Figura 5.11: Architettura sistema di messa in onda

L'unico elemento a cambiare è l'encoder, o meglio, il software che implementa la codifica; prima di vedere come questo debba essere modificato, vediamo come funziona il sistema.

5.2.2 Funzionamento

Mentre nella soluzione precedente si utilizzavano i metadati per segnalare l'inizio/-fine di uno spot, in questa soluzione vengono utilizzati gli ultrasuoni.

Gli ultrasuoni come visto nel capitolo introduttivo di questo documento, sono vibrazioni meccaniche della materia che si trasmettono sotto forma di segnali sonori la cui principale caratteristica consiste nel fatto che l'udito umano non le può captare perché hanno una frequenza superiore a 18 KHz.

Sfruttando questa caratteristica, sono stati inseriti in concomitanza dell'inizio e della fine di uno spot, dei segnali ad alte frequenze che non possono essere uditi dagli utenti ma che possono essere rilevati dall'applicazione android che riproduce l'audio. Prima di vedere quali sono questi segnali, vediamo da dove nascono e da quali tecnologie esistenti prendono forma.

Nel paragrafo 4.2 sono stati descritti i segnali DTMF che vengono utilizzati come sistema di segnalazione per le chiamate. Questi sono la composizione di due frequenze prestabilite e inconfondibili. Allo stesso modo si è provato a comporre nuovi segnali composti da una frequenza molto bassa (3 Hz) ed una molto alta (18 KHz) in modo che siano inudibili rispetto ai DTMF. Questa soluzione si è rivelata funzionante in un architettura semplice in cui l'audio non viene processato. Nelle architetture reali come quella di figura 5.1 questi segnali così composti non possono funzionare perché il processore audio presenta in ingresso un filtro passa alto che taglia tutte le frequenze inferiori ai 60 Hz in modo da eliminare tutti i disturbi provenienti dalla rete di alimentazione. Questo filtro va dunque ad eliminare la frequenza bassa (3Hz) del segnale precedentemente composto e lo rende inutilizzabile.

L'alternativa studiata è stata quella di generare dei segnali composti da frequenze entrambe molto alte, per esempio 18000 Hz e 18353 Hz, purtroppo questa soluzione non si è rilevata adottabile poiché le frequenze alte seppur inudibili prese singolarmente, andando a sommarle generano delle armoniche che possono essere udibili dall'orecchio umano.

A questo punto, si è cambiato approccio al problema, invece di generare segnali che sono la composizione di due frequenze, si è scelto di utilizzare segnali composti da più frequenze successive nel tempo.

In questo modo è stato possibile utilizzare frequenze molto alte e inudibili in modo singolo.

18000 Hz	18121 Hz	18223 Hz	18353 Hz
-----------------	-----------------	-----------------	-----------------

Tabella 5.1: Frequenze scelte in Hz

La tabella 5.1 mostra le frequenze scelte per generare i vari segnali. Queste sono maggiori o uguali di 18 Khz, in modo da essere inudibili, e abbastanza distanti tra di loro, in modo da essere inconfondibili.

Tramite queste frequenze è stato possibile generare dei segnali della durata di 1 secondo che sono la composizione delle quattro frequenze viste in tabella ognuna della durata di 250 millisecondi 5.1. In questo modo, modificando opportunamente la posizione di una frequenza rispetto alle altre, è possibile creare 16 combinazioni diverse.

1	18000	18121	18223	18353
2	18000	18121	18353	18223
3	18000	18353	18121	18223
4	18000	18353	18223	18121
5	18121	18000	18223	18353
6	18121	18000	18353	18223
7	18121	18353	18000	18223
8	18121	18353	18223	18000
9	18223	18000	18121	18353
10	18223	18000	18353	18121
11	18223	18353	18000	18121
12	18223	18353	18121	18000
13	18353	18000	18121	18223
14	18353	18000	18223	18121
15	18353	18223	18000	18121
16	18353	18223	18121	18000

Tabella 5.2: Segnali generati a partire dalle frequenze scelte

La tabella 5.2 mostra i 16 segnali creati a partire dalle 4 frequenze scelte. Al fine di questa soluzione, i segnali che servono sono solo due ma all'occorrenza, il numero dei segnali può crescere scegliendo nuove frequenze.

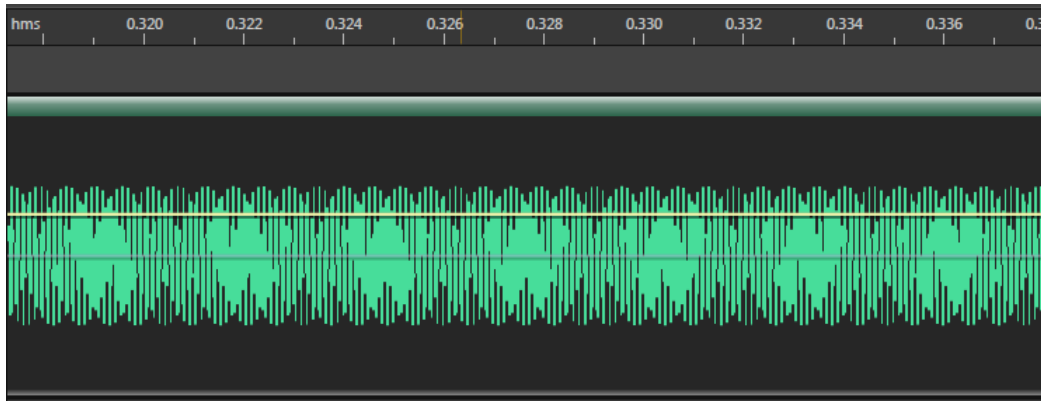


Figura 5.12: Forma d'onda di un segnale a 18 khz

La figura 5.12 mostra lo spettro di frequenza di uno dei segnali. Prendendo due di questi segnali, uno per segnalare l'inizio di uno spot e l'altro per la fine e inserendoli opportunamente all'interno del palinsesto avremo così implementato il meccanismo di segnalazione. I segnali vengono inseriti in concomitanza dell'inizio e della fine di uno spot, più precisamente nell'ultimo secondo di streaming prima dell'inizio della pubblicità e nel secondo finale dello spot. Di solito nei palinsesti radio viene inserito un jingle prima dell'inizio di uno spot pubblicitario, il segnale d'inizio può quindi essere inserito nel secondo finale del jingle.

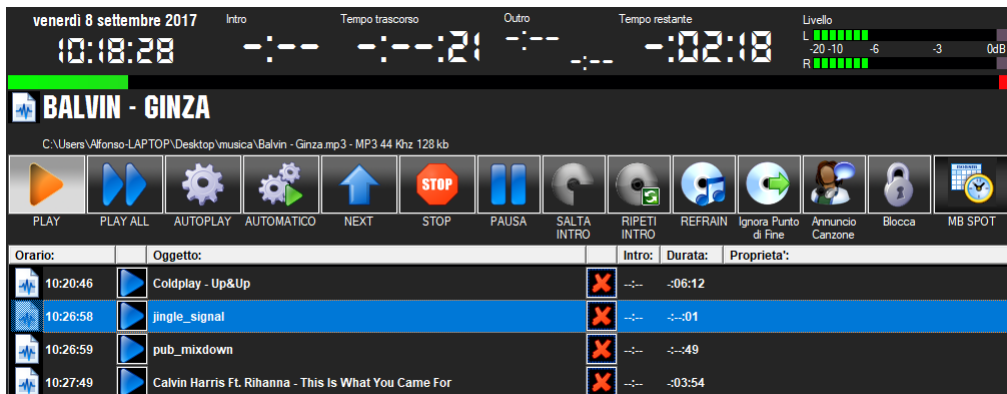


Figura 5.13: Palinsesto con jingle/segnale d'inizio spot

La figura 5.13 mostra un esempio di palinsesto che include un jingle con segnale che precede uno spot pubblicitario. Dentro il jingle è presente il segnale d'inizio spot ed è formato come mostrato nella figura 5.14.

Allo stesso modo, il segnale di fine spot deve essere inserito nel secondo finale dello spot pubblicitario.

Come detto nel paragrafo 5.2.1 l'unica modifica alla normale architettura radio, sta nell'encoder. Nel paragrafo 4.4 si è spiegato brevemente e in linea di principio come funziona un encoder audio, focalizzando l'attenzione sul fatto che gli encoder per migliorare la compressione, sfruttano dei teoremi della psicoacustica e tagliano le alte frequenze, non udibili dall'orecchio umano, dai brani originali. Il paragrafo poneva attenzione anche sul fatto che ad un maggior bit-rate di codifica, corrisponde una migliore qualità audio e quindi un minor taglio delle frequenze alte. Questa caratteristica gioca a sfavore di questa soluzione, infatti, le radio trasmettono

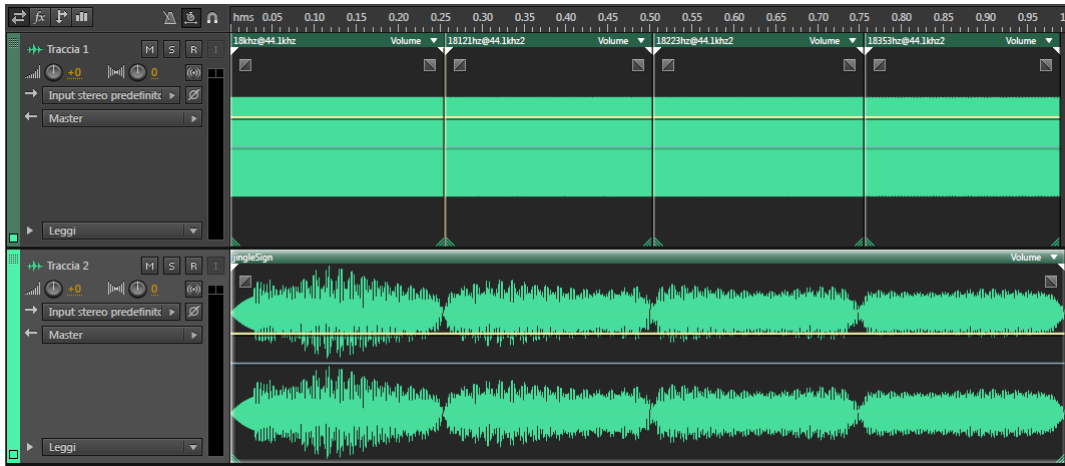


Figura 5.14: Forma d'onda del segnale composto con gli ultrasuoni

a bitrate di 64 Kbit/s o di 128 kbit/s, ed a questi livelli di bit-rate gli encoder tagliano le frequenze superiori ai 16 Khz e quindi anche tutti i segnali precedentemente descritti.

La soluzione più semplice a questo problema è quella di aumentare il bit-rate di trasmissione fino a 224 Kbit/s in modo che le frequenze nell'intorno dei 18 Khz non siano tagliate. Questa soluzione però non è adottabile dalle emittenti radiofoniche in quanto un aumento del bit-rate comporta un aumento della banda utilizzata per trasmettere. L'alternativa è quella di non modificare il bit-rate ma di andare a modificare l'encoder in modo che non tagli le frequenze di nostro interesse.

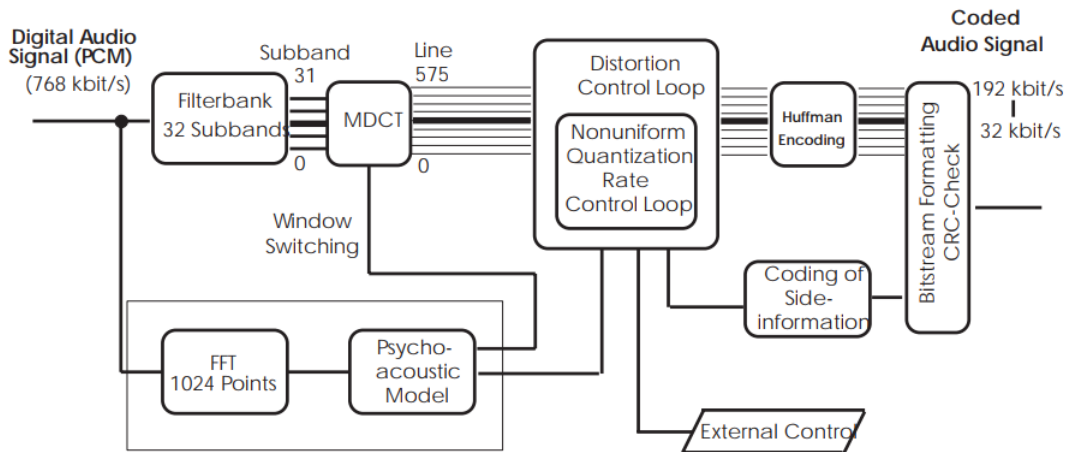


Figura 5.15: Diagramma a blocchi di un MPEG-1 Layer-3 encoder.

La figura 5.15 ripropone il modello seguito da un encoder mp3. Al fine di far passare le frequenze nell'intorno dei 18 Khz si deve agire sulla psicoacustica dell'algoritmo in modo che assegni un certo numero di bit per le frequenze di nostro interesse. Questa soluzione, seppur complessa dal punto di vista implementativo per via delle modifiche da apportare al codice sorgente dell'encoder, risulta quella più accreditabile da parte delle radio. Infine, lo stream codificato viene inviato ad un server di streaming, sia esso Wowza o Shoutcast che si occupa della distribuzione così come fatto nella soluzione con i metadati.

Ricapitolando, il software di regia, presenta degli elementi del palinsesto, modificati esternamente, che includono i segnali di inizio/fine spot; questi segnali, a differenza della soluzione con i metadati passa nello stesso canale dello streaming e non necessita di un canale secondario. Il segnale quando raggiunge l'encoder, non viene eliminato a patto che sia utilizzato un bit-rate maggiore o uguale a 224 Kbit/s o un encoder modificato per questa soluzione. Il flusso codificato raggiunge lo streaming server che lo distribuisce in rete.

5.2.3 Applicazione android

Come fatto nella soluzione precedente, vediamo ora come è strutturata l'applicazione, per dispositivi Android, in grado di effettuare la riproduzione del flusso audio e di effettuare la sostituzione della pubblicità quando segnalato dagli ultrasuoni.

A differenza dell'altro prototipo, per questa versione è stato scelto un metodo implicito (vedi par. 2.8.1) per la profilazione dell'utente. L'applicazione è accessibile tramite login social (Facebook o Google). Tramite il login social è infatti possibile recuperare alcune informazioni come l'età, il genere e la città di residenza da cui tracciarne un profilo. La figura 5.16 mostra la user interface per il login.

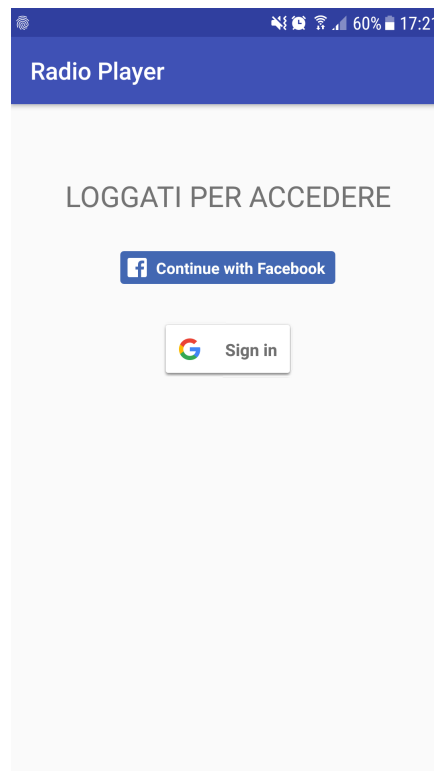


Figura 5.16: User interface per il login social.

Una volta effettuato il login, l'applicazione è in grado di tracciare il gruppo d'appartenenza adatto all'utente e di scaricare tramite richiesta http la lista degli spot pubblicitari adatti al profilo memorizzandoli sulla memoria locale del dispositivo in cui è in esecuzione l'applicazione.

L'activity principale consiste nei comandi per il player, ritroviamo, come nella soluzione precedente, i tre pulsanti di play, pause e live.

Anche questa soluzione fa uso di due player, uno per lo streaming principale e uno per la riproduzione degli spot personalizzati ma a differenza della soluzione con i metadati è stato utilizzato il framework multimediale Android MediaPlayer [5] che include il supporto per la riproduzione di vari tipi di supporti comuni, in modo da integrare facilmente audio, video e immagini nelle applicazioni per la riproduzione degli spot personalizzati e una soluzione più a basso livello, android Audiotrack [4], per la riproduzione dello streaming radio.

L'utilizzo di un player di basso livello come Audiotrack si è rilevato necessario in quanto bisogna riprodurre lo streaming di un buffer audio PCM e non dello streaming codificato che viene ricevuto dal server. Infatti, per implementare il sistema di segnalazione, è necessario che il flusso streaming sia decodificato manualmente per ottenere il buffer PCM da analizzare. Al fine di ottenere lo streaming decodificato in PCM, l'applicazione implementa un decoder mp3, mpg123 [20], completamente scritto in C++ al fine di ottimizzare la velocità di decodifica. Android permette tramite un kit di sviluppo nativo (NDK) di utilizzare codice C e C++ [7].

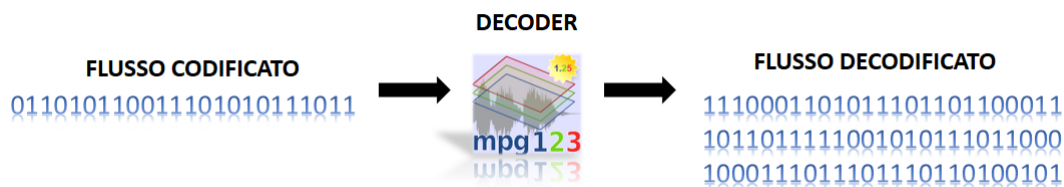


Figura 5.17: Rappresentazione del flusso di decodifica.

Il flusso decodificato viene analizzato tramite l'algoritmo di Goertzel (vedi par. 3.4) che è stato precedentemente configurato in modo da rilevare le frequenze d'interesse. Se l'algoritmo rileva una frequenza d'interesse questa viene inviata ad un oggetto chiamato SequenceChecker che verifica se la sequenza è una sequenza valida oppure no.

Il SequenceChecker implementa una macchina a stati con 4 stadi, inizio sequenza, intermedio uno, intermedio due e fine sequenza. Quando il Goertzel rileva una frequenza, questa viene inviata al SequenceChecker insieme al tempo in millisecondi di rilevamento; se la macchina a stati è sullo stato d'inizio, la frequenza viene memorizzata come frequenza uno e si passa allo stato di intermedio uno, se è nello stato di intermedio uno viene memorizzata come frequenza due e si passa allo stato di intermedio due, se è su intermedio due viene memorizzata come frequenza tre e si passa allo stato di fine sequenza, se è su fine sequenza viene memorizzata come frequenza quattro e viene verificata che la sequenza sia una frequenza valida, se la frequenza è valida viene innescato il meccanismo di crossfading, che come nella soluzione con i metadati, abbassa il volume del player principale e fa partire il player con la pubblicità personalizzata. Negli stati intermedi, se il tempo di rilevamento della frequenza è maggiore di 250 millisecondi rispetto alla frequenza dello stato precedente, la sequenza viene resettata e si passa allo stato di inizio sequenza. Con questo accorgimento si vuole aumentare l'affidabilità dell'algoritmo in modo che non ci possano essere delle "spurious frequencies" che creano sequenze valide e conseguentemente innescano il sistema di crossfading.

La figura 5.18 illustra la macchina a stati del SequenceChecker, mentre la figura 5.19 mostra la user interface del player principale Audiotrack [4]. Quest'ultimo viene

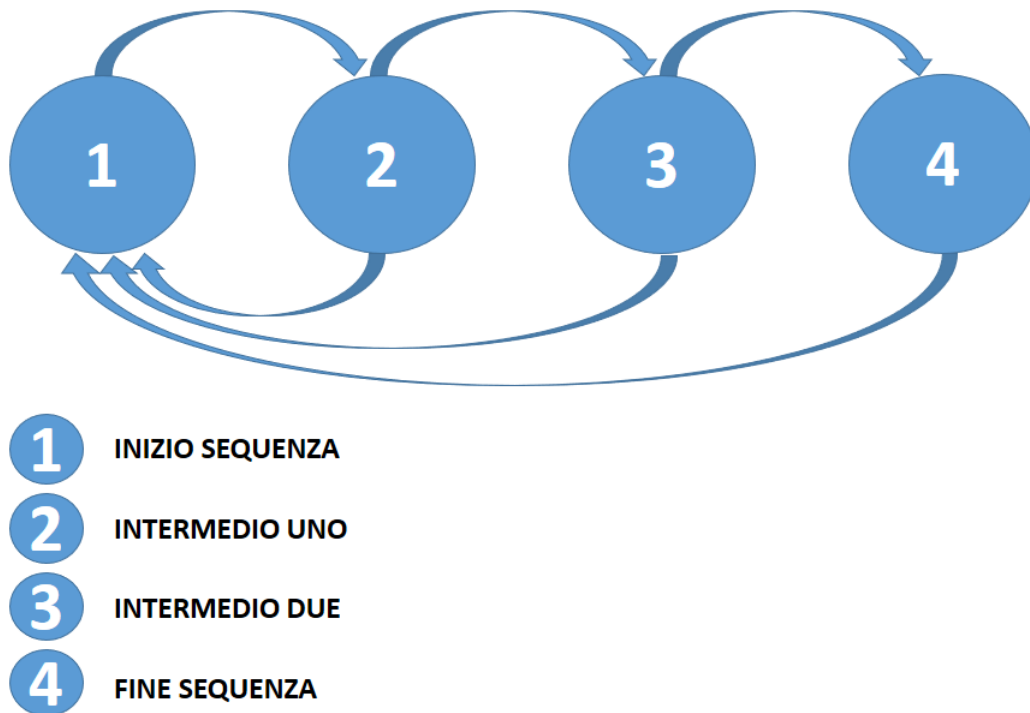


Figura 5.18: Macchina a stati del SequenceChecker.

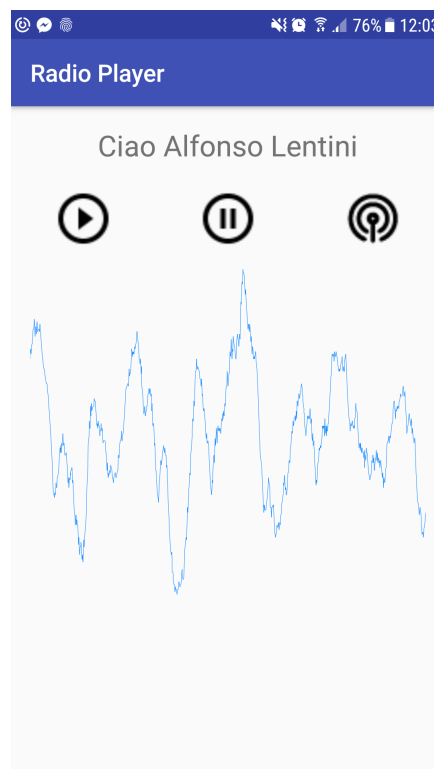


Figura 5.19: User interface del player principale.

gestito dallo stesso thread che si occupa dello streaming e della decodifica in modo da mantenere tutto il flusso sincrono, mentre il player degli spot personalizzati, è

gestito da un android Service[6].

Per riassumere, l'applicazione riceve il flusso streaming della web radio codificato; tramite una libreria esterna questo viene decodificato ottenendo un segnale PCM in modo da poter essere analizzato dall'algoritmo di Goertzel. Il Goertzel analizza tutto il flusso e quando rivela una delle frequenze scelte (vedi tab. 5.1) la passa al SequenceChecker che implementa una macchina a stati. Se la macchina a stati arriva al quarto stato con una sequenza valida allora viene attivato il sistema di crossfading dei player.

Il sistema di crossfading è essenzialmente uguale a quello della soluzione con i metadati, consiste nell'innalzare il volume del player che deve riprodurre lo spot pubblicitario, mentre si abbassa il volume del player dello spot principale quando viene segnalato uno spot. Allo stesso modo quando viene segnalata la fine di uno spot, viene fatto un crossfading inverso con il player principale che aumenta di volume e il player degli spot personalizzati che lo diminuisce.

Infine, nel prossimo capitolo vengono presentati i risultati ottenuti con le due soluzioni proposte in questo capitolo.

Capitolo 6

Conclusioni

In questo documento sono state presentate tutte le fasi dalla progettazione all'implementazione di un sistema di segnalazione al fine di sostituire inserzioni pubblicitarie all'interno del flusso streaming di una web radio. Sono state descritte le motivazioni dietro allo sviluppo di queste tecnologie, i problemi che stavano alla base, lo stato dell'arte, le tecnologie/algoritmi utilizzati passando infine alle scelte progettuali e all'implementazione di due sistemi di segnalazione. In quest'ultimo capitolo vengono presentati i risultati, evidenziando i pro e i contro ottenuti nelle due soluzioni precedentemente proposte nel capitolo 5. Inoltre vengono evidenziati i possibili sviluppi futuri che possono prendere vita tramite l'utilizzo di queste tecnologie.

6.1 Risultati, vantaggi e svantaggi

Le soluzioni presentate nel capitolo 5 sono entrambe valide ma presentano caratteristiche e adottabilità diverse. Ambedue le soluzioni introducono un lavoro da fare da parte del regista della radio; come visto, sia per i metadati che per gli ultrasuoni bisogna effettuare una modifica al palinsesto tramite il programma di regia. Questa modifica consiste nell'impostazione dei bit corretti sui punti del palinsesto in modo da inviare i corretti segnali alla scheda a contatti per l'iniezione del metadato e per la seconda soluzione, nell'elaborazione, esterna al programma di regia, degli spot pubblicitari in modo da sommare alla traccia audio i segnali ad ultrasuoni.

In entrambi i casi questo può essere uno svantaggio, il regista deve essere istruito sul sistema e su come impostare gli spot in modo che il sistema funzioni correttamente. Proseguendo nell'architettura dei sistemi, abbiamo visto che la soluzione con i metadati, necessita l'introduzione di due componenti che di solito non si trovano in una comune architettura radio, la scheda a contatti e il software di iniezione dei metadati. Questo è un possibile svantaggio di questa soluzione poiché prevede una piccola modifica all'architettura che però non è critica in quanto i costi e la complessità dell'operazione non è poi così tanto onerosa.

D'altro canto la soluzione ad ultrasuoni non prevede l'introduzione di nuovi componenti ma la modifica di quelli già esistenti. Infatti, come detto ampiamente nei capitoli precedenti, esiste un problema cruciale sull'encoder responsabile dell'eliminazione delle alte frequenze introdotte. La soluzione presentata consiste nella modifica dell'encoder al fine di cambiare il comportamento del modello di psicoacustica in modo da lasciar passare le frequenze utilizzate per generare i segnali. Questa è la soluzione più accreditabile in quanto l'altra possibilità prevede l'innalzamento

del bit-rate di codifica, che tendenzialmente si attesta tra i 64 e i 128 kbit/s, alla più onerosa frequenza di 224 kbit/s. Quest'ultima possibilità seppur la più semplice dal punto di vista implementativo, potrebbe venire scartata dalla maggior parte delle emittenti radio in quanto andrebbe ad aumentare sostanzialmente la banda d'uscita del flusso streaming ed inoltre andrebbe ad aumentare il traffico dati utilizzato dall'utente.

Per il resto, l'una e l'altra soluzione non presentano modifiche invasive all'architettura ne introducono lavori extra da parte delle emittenti radiofoniche, le uniche variazioni sono quelle appena elencate.

Dal punto di vista funzionale, le due soluzioni, seppur diverse nel loro essere, offrono le stesse prestazioni, infatti riescono ad assicurare un errore massimo di 100 ms.

6.2 Sviluppi futuri

6.2.1 Podcast

Nel mondo delle radio così come già accaduto nella televisione, è nato il concetto di "on demand". Per on demand si intende l'accesso alle risorse informatiche solo quando necessario, eventualmente pagando le stesse in base all'utilizzo e non in base a un canone fisso o acquistando una licenza una tantum. Il VoD - Video on Demand è un servizio interattivo della TV di ultima generazione che permette di guardare un programma su richiesta in qualsiasi momento. Il servizio è interattivo proprio perché è lo spettatore che si può creare un palinsesto altamente personalizzato di TV on Demand (su domanda o richiesta). Chiaramente la scelta dei programmi e dei contenuti da richiedere è limitata dalle registrazioni a disposizione dell'emittente se si tratta di TV che trasmettono in digitale terrestre o satellitare.

Il podcasting è l'insieme delle tecnologie e delle operazioni relative al download automatico di file di qualsivoglia natura (detti podcast, fusione e contrazione di PoD, "Personal on Demand"), tramite un'infrastruttura di trasmissione dati e un programma client chiamato "aggregatore" o feed reader.

Iscriversi ad un podcast permette all'utente di ottenere file riproducibili anche offline e di disporre di una grande quantità di fonti da cui attingere. Al contrario il broadcast offre una sola trasmissione alla volta e obbliga ad essere sintonizzati ad una determinata ora. L'ascolto di audio in streaming su internet può eliminare l'obbligo di sintonizzazione in un determinato momento dato dalle trasmissioni tradizionali (come accade nel caso del video/audio on demand), tuttavia offre comunque una sola risorsa alla volta e obbliga l'utente a essere connesso a Internet durante la riproduzione del file. La capacità di ricevere automaticamente pubblicazioni da fonti multiple è proprio uno dei punti di forza che distinguono il podcasting dalle trasmissioni sia tradizionali sia in streaming. Per permettere la registrazione di un contenuto radiofonico, risulta necessario poter marcare questi contenuti all'interno del palinsesto. In questa ottica si prestano bene le soluzioni utilizzate per la sostituzione delle pubblicità. Allo stesso modo di quanto fatto per le inserzioni pubblicitarie, tramite un metadato o un segnale ad ultrasuoni si potrebbe marcare l'inizio e la fine dei vari programmi in modo da permetterne l'individuazione e la registrazione.

6.2.2 Sincronizzazione in-door

L'implementazione di un sistema di comunicazione basato sull'uso di segnali ad ultrasuoni apre numerosi scenari d'applicazione. Uno dei campi in cui è possibile applicare questa tecnologia è quello della sincronizzazione in-door. A tale scopo è stato pensato un sistema in grado di creare esperienze di sincronizzazione e navigazione all'interno di musei.

L'implementazione è fondata sull'utilizzo di micro-controllori posizionati debitamente in grado di generare segnali ad ultrasuoni che possano essere rilevati da apposite applicazioni mobile. Un esempio di realizzazione di questi diffusori è dato dall'impiego di NodeMCU, una piattaforma open source sviluppata per l'IoT che include un modulo Wi-Fi ESP8266.

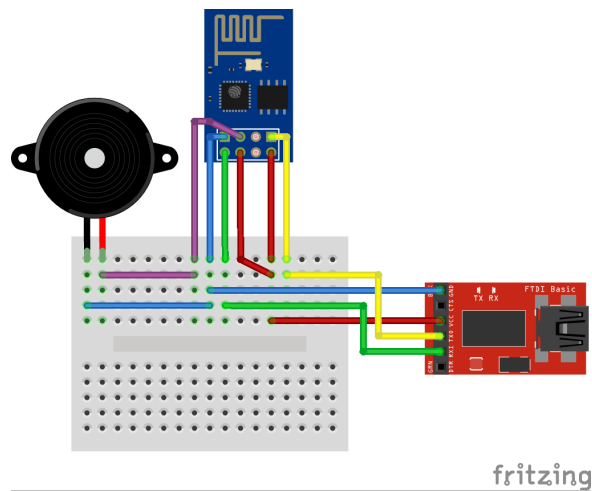


Figura 6.1: NodeMcu con amplificatore e speaker ad ultrasuoni

Tramite l'utilizzo del modulo wifi è possibile pilotare questi diffusori in modo che generino segnali diversi a secondo del posto in cui sono stati posizionati. L'utente in questo modo è in grado ricevere informazioni che lo indirizzino o lo istruiscano in base alla posizione in cui si trova.

Questo progetto seppur ancora in fase embrionale, ha tutte le potenzialità per essere utilizzato in tutti quei posti dove le tecnologie di comunicazione moderne, per vari motivi, non possono essere utilizzate.

Bibliografia

- [1] Soulodre G. A. et al. “Subjective evaluation of state-of-the-art 2-channel audio codecs”. In: *J. Audio Eng. Soc.*, 46(3):164 – 176 (1998).
- [2] D. Allen. “Journal of Radio Studies”. In: (nov. 2007).
- [3] Android. *Android Activity*. URL: <https://developer.android.com/reference/android/app/Activity.html>.
- [4] Android. *Android AudioTrack*. URL: <https://developer.android.com/reference/android/media/AudioTrack.html>.
- [5] Android. *Android MediaPlayer*. URL: <https://developer.android.com/guide/topics/media/midiplayer.html>.
- [6] Android. *Android Service*. URL: <https://developer.android.com/guide/components/services.html>.
- [7] Android. *Native Development Kit (NDK)*. URL: <https://developer.android.com/ndk/index.html>.
- [8] G. Araniti et al. “Adaptive controlling the QoS of multimedia wireless applications through user profiling techniques”. In: *IEEE Journal on selected areas in communication*, 21(10), pp. 1546-1556 (2003).
- [9] Barry Berman. “Keeping DJ Endorsements Alive”. In: *MediaWeek* (mar. 2008).
- [10] K. Brandenburg e Marina Bosi. “Overview of MPEG audio: Current and future standards for low bit-rate audio coding.” In: *J. Audio Eng. Soc.*, 45(1/2):4 –21 (1997).
- [11] K. Brandenburg e G. Stoll. “ISO-MPEG-1 Audio: a generic standard for coding of high quality digital audio”. In: *N. Gilchrist and Ch. Grewin, editors, Collected Papers on Digital Audio Bit-Rate Reduction, pages 31 – 42* (1996).
- [12] *Broadband TV News*. URL: <http://www.broadbandtvnews.com/2014/10/16/kantar-buys-civolution-audio-watermarking-unit/>.
- [13] M. A. De Jesús et al. “Nonuniform Discrete Short-Time Fourier Transform A Goertzel Filter Bank versus a FIR Filtering Approach”. In: *IEEE International Midwest Symposium on Circuits and Systems, Vol. 2, pp. 188-192* (2006).
- [14] Forrester. *In-Stream Audio Advertising Is Rife with Marketing Opportunity*. URL: <https://www.forrester.com/InStream+Audio+Advertising+Is+Rife+With+Marketing+Opportunity/-/E-PRE5764>.
- [15] Elizabeth Gasiorowski-Denis. *Adobe Extensible Metadata Platform (XMP) becomes an ISO standard*. 2012. URL: http://www.iso.org/iso/home/news_index/news_archive/news.htm?refid=Ref1525.

- [16] C. Gena. “Methods and techniques for the evaluation of user-adaptive systems”. In: *The Knowledge Engineering*, 20(1), pp.1-37 (2005).
- [17] A. Godoy D. Amandi. “User profiling in personal information agents: a survey”. In: *The Knowledge Engineering Review Journal*, 20(4), pp. 329-361. (2005).
- [18] Google. *ExoPlayer*. URL: <https://developer.android.com/guide/topics/media/exoplayer.html>.
- [19] Google. *Interactive Media Ads*. URL: <https://developers.google.com/interactive-media-ads/>.
- [20] Michael Hipp e Thomas Orgis. *mpg123*. URL: <https://www.mpg123.de/index.shtml>.
- [21] Brandenburg Karlheinz. “MP3 and AAC Explained”. In: *Fraunhofer Institute for Integrated Circuits FhG-IIS A, Erlangen, Germany* ().
- [22] Charan Langton e Victor Levin. “Intuitive Guide to Fourier Analysis”. In: *Complex to Real* ().
- [23] MPEG. “Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s, part 3: Audio. International Standard IS 11172-3, ISO/IEC JTC1/SC29 WG11”. In: (1992).
- [24] RAIN News. *Streaming radio is the most mobile content category*. URL: <http://bit.ly/1iKbLVB>.
- [25] Nielsen. *Nielsen Audio*. URL: <http://www.nielsen.com/audio>.
- [26] Justin Nielson. “New SNL Kagan Internet music study: 10-year revenue growth; music licensing fees a ‘burden’”. In: (nov. 2014).
- [27] M. Nilsson. *Id3v2 - Informal Standard*. 1999. URL: <http://id3.org/id3v2.3.0>.
- [28] D. Poo, B. Chng e J. M. Goh. “A hybrid approach for user profiling”. In: *Annual Hawaii International Conference on System Sciences*, 4(4), pp. 1-9 (2003).
- [29] David Porter. *Why Internet Radio is The Biggest Advertising Opportunity of The Future*. 2015. URL: <https://www.forbes.com/sites/davidporter/2015/03/18/why-internet-radio-is-the-biggest-advertising-opportunity-of-the-future/#51b6115171af>.
- [30] Mb Radio. *Mb Studio*. URL: <http://www.mbradio.it>.
- [31] Henczel S. “Creating user profiles to improve information quality”. In: *Factiva*, 28(3), p. 30 (2004).
- [32] Wowza Media Server. *Wowza Streaming Engine - Informal Standard*. URL: <https://www.wowza.com/docs/wowza-streaming-engine-software-updates>.
- [33] Peter D. Welch. “The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms”. In: *IEEE Transactions on audio and Electroacoustics*, Vol. AU-15, No.2 (1967).

- [34] David B. Wilkerson. “Clear Channel: Ad Cuts a Success”. In: *Marketwatch* (feb. 2005).
- [35] Roy Williams. “Radio Ads: How Long Should They Be?” In: (set. 2014).
- [36] Filip Zápłata e Miroslav Kasal. “Using the Goertzel algorithm as a filter”. In: *Radioelektronika (RADIOELEKTRONIKA), 2014 24th International Conference* (2014).