

Perizia Fonica con comparazione del parlante

Date : 16 febbraio 2017



Perizia fonica forense: concetti fondamentali

Per poter redigere un'accurata e completa perizia fonica con comparazione vocale bisogna innanzitutto conoscere approfonditamente alcuni concetti fondamentali per effettuare le analisi necessarie e rispondere ai quesiti posti dal committente.

La quasi totalità dei suoni che udiamo nel mondo reale non sono suoni semplici, o puri, ma suoni complessi, cioè composti da una serie di suoni puri.

Identificativi del suono sono frequenza ed ampiezza:

- La frequenza, indicata in Hertz, indica il numero di oscillazioni (cicli) compiute in un secondo.
Lo spettro udibile umano è compreso tra i 20 Hz ed i 20.000 Hz, maggiore sarà la frequenza maggiore sarà la nostra percezione dell'altezza del suono;
- L'ampiezza indica l'energia di una frequenza, possiamo per semplicità di comprensione paragonarla al volume.

La scomposizione di un suono complesso in una somma algebrica di suoni puri avviene attraverso una trasformata integrale nota con il nome di "trasformata di Fourier", dal fisico francese Jean Baptiste Joseph Fourier.

La trasformata di Fourier consente di analizzare il contenuto dello spettro dell'udibile in modo numerico, attribuendo così un valore energetico preciso per ogni frequenza (o meglio banda di frequenze) presente nel programma preso in analisi. La versione semplificata dell'analisi di Fourier utilizzata nei software di analisi acustica è chiamata Fast Fourier Transform (FFT).

Per meglio comprendere la somma di frequenze pure utilizzeremo un'analogia con l'ottica. È noto come alcuni colori, detti fondamentali, siano puri, cioè non ulteriormente scomponibili. Questi colori sono il rosso, il giallo ed il blu. A ciascuno di essi corrisponde una certa lunghezza d'onda del raggio luminoso ed il prisma (che scompone la luce bianca nei sette colori dello

spettro luminoso) mostrerà solo quella componente.

Il medesimo processo avviene per il suono: ad una certa lunghezza d'onda del suono corrisponde una certa altezza percepita. Se non è presente contemporaneamente nessun'altra frequenza, il suono sarà puro. In sintesi, due delle grandezze fondamentali che ci consentono di analizzare oggettivamente un suono sono frequenza ed ampiezza.

La traduzione digitale ideale dei segnali audio è sottoposta ad alcune limitazioni, oltre ad avere diversi pregi: la banda passante (il range delle frequenze prese in considerazione) di una registrazione è sottoposta alla legge di Nyquist, che vincola la frequenza maggiore contenuta nel segnale alla frequenza di campionamento nell'ordine di 2:1.

Per campionare correttamente (tradurre in digitale) un segnale analogico è necessario che la frequenza di campionamento sia almeno il doppio della frequenza massima contenuta nel programma (per un segnale a 5000 Hz è necessaria una frequenza di campionamento di 10.000 Hz). La traduzione digitale dell'ampiezza è invece data dalla quantità di bit. Per 16 bit (il numero di bit contenuti in un cd musicale, ad esempio) ci sono 65535 possibili livelli di ampiezza. Maggiore sarà il numero di bit, più alta la frequenza di campionamento, migliore sarà la traduzione in dominio digitale del programma audio.

LE CORDE VOCALI

Senza entrare in particolari fisiologici, le corde vocali sono due lamelle che vengono messe in vibrazione dal passaggio dell'aria. Non essendo particolarmente morbide, la loro vibrazione è di tipo impulsivo, producendo una serie di "toc" molto ravvicinati seguiti da uno spazio vuoto. Tale spazio viene poi riempito dalla risonanza del tratto vocale.

All'impulso iniziale segue una risonanza che, in pratica, è una piccola serie di echi generati dal tratto vocale che agisce come cassa armonica. Ciò significa che, nei suoni tenuti, le corde vocali, da sole, non sono in grado di produrre timbri diversi, ma solo altezze diverse.

Bisogna ricordare, poi, che non sempre le corde vocali entrano in azione, infatti la maggior parte delle consonanti è prodotta mediante rumori generati dal passaggio dell'aria in un tratto vocale opportunamente configurato dalla posizione della mandibola, della lingua e delle labbra.

ANALISI DELLE VOCI

L'onda prodotta dalle vibrazioni delle corde sonore ha una struttura composta dalla frequenza fondamentale [F0] dell'onda vocale corrisponde al numero di oscillazioni delle corde vocaliche nell'unità di tempo. L'onda prodotta dalle corde vocali, prima di venir modificata dal tratto vocalico, viene chiamata anche onda glottidale.

La frequenza fondamentale è percepita come tono della voce [timbro vocalico]. Il valore di [F0] dipende dalle caratteristiche anatomiche del parlante (es. spessore e lunghezza delle corde vocaliche) e dalla pressione ipoglottidale. Più spesse e lunghe sono le pliche, più bassa è la

frequenza di vibrazione, difatti il tono della voce di un bambino è più alto di quello di una donna e quella di una donna è più alto di quella di un uomo.

Come viene analizzato il segnale vocale?

Nel parlato lo spettro cambia continuamente, quindi c'è bisogno di visualizzare i cambiamenti dello spettro nel tempo, a questo scopo si usano gli spettrogrammi.

Negli spettrogrammi, la linea orizzontale rappresenta il tempo da sinistra a destra e la linea verticale rappresenta gli aumenti di frequenza, dal basso verso l'alto. L'ampiezza è una terza dimensione, dovrebbe essere rappresentata su un piano perpendicolare al piano dell'asse del tempo delle frequenze sull'asse delle x, ma per semplificare gli spettrogrammi è stata adottata la convenzione di rappresentare l'ampiezza con scale di grigio o di colore [più scuro = più ampio].

Con gli spettrogrammi si possono definire:

- Durata vocalica e consonantica;
- Timbro delle vocali;
- Epentesi vocalica e consonantica;
- Variazione intra e interlinguistica.

I SUONI VOCALIZZATI

Intanto bisogna ricordare che le vocali sono ben più di 5. Nelle varie forme aperte, chiuse, mezze-aperte, mezze-chiuse, la lingua inglese ne conta da 10 a 14. L'italiano ha un doppio accento solo sulla 'e', ma, se si considerano anche i dialetti, si arriva facilmente alla decina.

La generazione delle vocali avviene nel modo seguente: il tratto vocale, che va dalle corde fino alle labbra, costituisce una vera e propria cassa armonica che filtra il suono prodotto dalle corde generando fino a 5 formanti, caratterizzando in modo pressoché definitivo il timbro di una voce, ovvero la distribuzione dell'energia (dovuta alle risonanze del tratto vocalico) durante la modulazione del parlato. Ogni suono vocalizzato ha una propria caratteristica di configurazione di formanti.

SAGGIO FONICO

Fase importantissima della perizia fonica forense di comparazione vocale è il saggio fonico, ovvero l'acquisizione della voce da comparare acquisita e registrata con strumentazione appropriata, come per esempio lo ZOOM Handy Recorder H6, operazione che deve essere effettuata in un ambiente consono, quindi in un luogo silenzioso e privo di interferenze acustiche indesiderate (come l'eco ad esempio).

Inizialmente viene effettuato un colloquio introduttivo nel quale viene richiesto al parlatore le generalità. Successivamente vengono fatte leggere delle frasi appositamente scelte sotto

dettatura e infine viene fatto un colloquio informale e redatto il verbale.

Durante il saggio fonico è molto importante mettere a loro agio la persona registrata, in modo che abbia la voce più chiara e naturale possibile.

TECNICHE E METODOLOGIE DI ANALISI



Analisi 1: Ascolto Soggettivo

L'ascolto soggettivo è un esame di ascolto estremamente scrupoloso, effettuato in condizioni ottimali, ovvero in studio professionale dove il rapporto segnale rumore è eccellente e l'intero percorso del segnale è mantenuto a standard di efficienza impeccabili.

Le caratteristiche e gli elementi di interesse che caratterizzano una voce sono: timbro, intonazione, posizione degli accenti, velocità di locuzione, durata ed epentesi vocalica e consonantica, variazione intra e interlinguistica.

KayPentax Multi-Speech 3700, Steinberg Cubase Elements 7, iZotope RX5 sono alcuni tra i migliori software disponibili su sistemi Windows.

L'intera catena di riproduzione è stata tenuta in ambiente digitale ed ottico, eliminando qualsiasi degenerazione del segnale originale. È necessario tener presente che la tecnica di ascolto di diversi professionisti, posti in condizioni ottimali, è un primo passaggio di analisi estremamente indicativo per la scelta delle metodologie di ricerca successive e per la selezione dei punti critici.

Analisi 2: Rappresentazione Grafica del Segnale (Waveform)

L'ascolto delle tracce repertate contestualmente alla visione in alta definizione della rappresentazione grafica del segnale, fornita dal software di analisi iZotope RX5, consente di

“vedere” in dettaglio alcune sezioni del file critiche e consente di visualizzare eventuali anomalie nella sequenza dei campioni registrati.

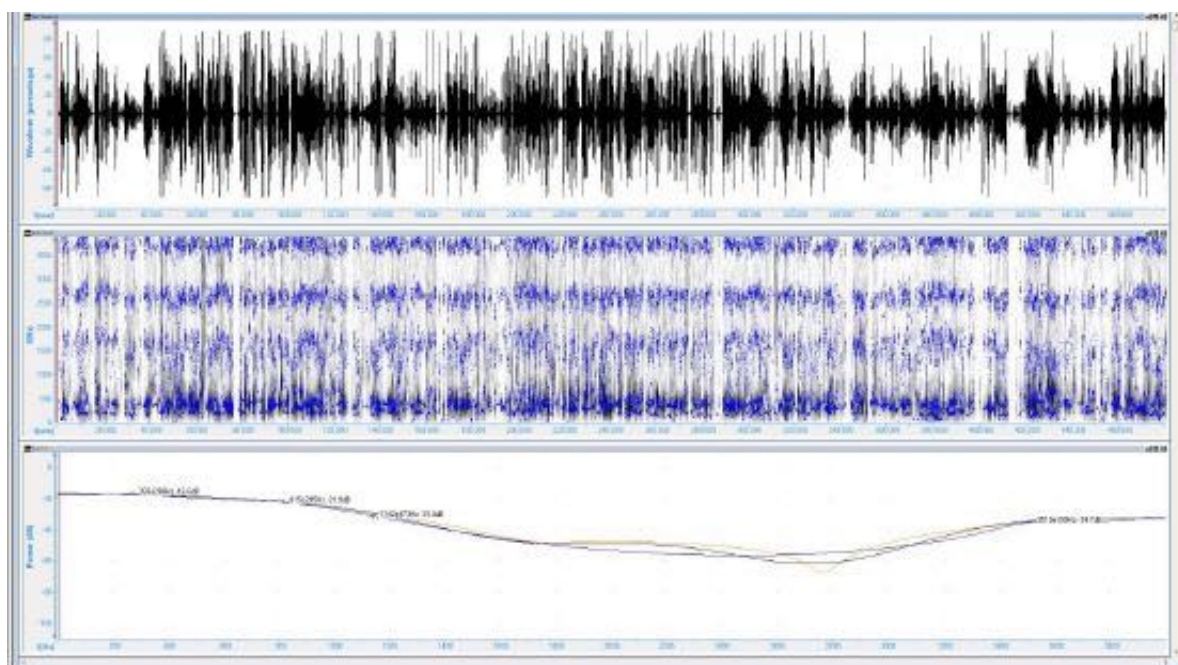
Analisi 3: Spettrogrammi (Spectrogram SPG)

Laddove nelle analisi precedenti si fossero verificati punti interessanti o non chiari si procede ad una analisi più approfondita: l'analisi spettrografica.

Lo spettrogramma è un sistema di analisi che consente di verificare il contenuto in frequenza e dinamica di un programma audio in ogni istante. Sull'asse delle ascisse è riportato il tempo, sull'asse delle ordinate la frequenza e l'intensità è tradotta graficamente con un aumento della densità dei punti. In questo modo è possibile vedere in ogni istante quali frequenze sono presenti e con quale intensità. Questo consente ad un tecnico esperto di “vedere” oltre che ascoltare il contenuto audio e confrontarlo con altri segnali.

Dove possibile si ricorre ad una visualizzazione colorata in base a scale cromatiche per rendere in modo chiaro, anche a non professionisti, la variazione di frequenza ed intensità in un dato momento. L'uso degli spettrogrammi è fondamentale nel riconoscimento vocale del parlatore, in quanto capace di scomporre il segnale audio complesso di una voce nelle sue frequenze fondamentali.

In funzione delle intensità di alcune bande di frequenza si può risalire a quelle che vengono definite formanti della voce, che sono peculiari e riconducibili con buona approssimazione ad un parlatore.



Analisi 4: Formanti (Formant FMT)

Come accennato nel punto precedente, le formanti di un segnale sono funzione delle intensità rilevate in alcune particolari bande di frequenza. L'uso dell'analisi delle formanti consente di ottenere una serie di valori grafico/numerici utili alla valutazione delle voci.

I parametri di analisi utilizzati, limitatamente a questo procedimento, sono stati scelti in funzione della rappresentazione di alcune voci specifiche. La procedura prevede un'accurata scelta del punto di analisi al quale applicare un LPC [Linear Predictive Coding]. LPC è una tecnica di analisi di segnali vocali, ideata per consentire una digitalizzazione di buona qualità con bassi valori di bitrate. Il principio di base sta nella assunzione secondo la quale la voce è il risultato dalla modulazione provocata da gola e bocca (detta formante) della emissione sonora da parte delle corde vocali (il residuo). Secondo tale assunzione, la formante può essere predetta mediante un'equazione lineare che tenga conto dei campioni precedenti ed il residuo dalla sottrazione della formante dal campione.

Generalmente un segnale vocale è composto da due componenti: un insieme di coefficienti per la predizione lineare e un errore di predizione. Dai risultati della procedura di analisi vengono estrapolati i valori delle formanti.

IDENTIFICAZIONE E RICONOSCIMENTO DEL PARLANTE

La valutazione e l'identificazione o riconoscimento di un parlante, soprattutto relativamente alle intercettazioni/registrazioni ambientali e telefoniche o di bassa qualità audio, non può essere espressa in modo definitivo.

Qualora ci fosse l'ausilio di un saggio fonico, ovvero una registrazione in alta definizione dell'imputato eseguita dal Perito, il margine di errore si ridurrebbe.

I limiti intrinseci dei reperti e le caratteristiche vocali permettono di ricondurre, quando possibile, ad un parlante con "buona probabilità", la certezza assoluta non è plausibile dato il numero di variabili presenti nella registrazione, come indicato nei paragrafi precedenti.

Dato determinante in relazione alle intercettazioni ambientali, è la valutazione dell'ambiente di registrazione: la quota di rumore deve essere ridotta al minimo possibile.

L'eventuale presenza di altre voci sovrapposte, macchine o motori, musica ed altri eventi tali da modificare in modo significativo lo spettro audio relativo al parlato, portano ad una drastica riduzione della probabilità di individuazione certa delle caratteristiche peculiari del parlante.

Il responso emerso da tutte le analisi descritte precedentemente viene inserito all'interno della perizia nella quale viene indicata la probabilità che la voce registrata durante il saggio fonico sia la stessa che si sente nel file audio oggetto di indagini.

A cura di: **Michele Vitiello**