



La raccolta di file audio per trascrizione automatica

Linee guida

Versione del documento.: 1.0

Ultimo aggiornamento: 2008-07-14

Indice

Introduzione.....3

Consigli generali.....3

 Cosa fare in caso di una registrazione di bassa qualità.....4

Riconoscere una buona registrazione.....4

 Specifiche tecniche.....4

 Le proprietà di un file audio.....4

Software utili.....7

 Audacity.....7

 dBPowerAmp.....8

L'ambiente di registrazione.....8

 Riverbero.....8

 Eco.....8

I tipi di microfono.....9

 Unidirezionale: cardiode.....9

 Unidirezionale: supercardioide9

 Omnidirezionale.....9

Le fonte di rumore10

I problemi legati al parlatore.....10

 Emotività del parlato spontaneo.....10

 Sovrapposizione di più parlatori.....11

 Parlato dialettale.....11

Ulteriori informazioni.....11



Introduzione

FlyScribe è un sistema automatizzato per la trascrizione da una registrazione audio in un file di testo. Per effettuare questa operazione è necessario partire da un audio di buona qualità. Arrivare alla qualità richiesta può sembrare facile. Però, per ottenere risultati sicuri ed efficaci ogni volta, è importante conoscere alcune regole di base.

Questo documento presenta suggerimenti che possono aiutare l'utente a migliorare la qualità delle acquisizioni audio. Questi accorgimenti portano naturalmente a migliorare la qualità dell'audio in sé ma sono anche utili alle trascrizioni automatiche di FlyScribe.

Infine, il mondo della registrazione audio è molto ampio e qui presentiamo soltanto pochi accorgimenti di base. In questo senso, è bene tenere presente che il miglioramento della registrazione è raggiungibile tramite tentativi e accettando che non sarà mai perfetto.

Consigli generali

Durante la registrazione è consigliato l'ascolto del suono mediante una cuffia, in quanto quello che noi percepiamo attraverso l'ambiente è diverso da quello che viene effettivamente rilevato dal microfono. Un esempio banale può essere quello di provare a registrare in un luogo silenzioso e quindi riascoltare. Durante la fase di riascolto si può notare che il risultato della registrazione non è affatto una situazione di silenzio: ci sono sempre rumori di fondo di diversa natura dovuti a vari fattori che non eravamo in grado di notare prima del riascolto.

Un grosso rischio è quello di compromettere inutilmente una situazione già buona, quindi prima di modificare le impostazioni relative alla fase di registrazione è opportuno verificare la qualità delle registrazioni ottenute nelle condizioni attuali.

Nel caso in cui la registrazione presenti un parlato di buona qualità ma sia presente del rumore, è consigliato provare ad abbassare il volume del microfono e chiedere al parlatore di provare ad alzare un po' la voce. Inoltre se il microfono è orientato verso una fonte di rumore il consiglio è quello di spegnere (se possibile) la fonte del rumore o di cercare di riorientare il microfono.

Se il volume della voce tende ad essere troppo alto, un buon accorgimento può essere quello di alzare il volume dell'impianto in sala cosicché il parlatore sia indotto ad abbassare il tono della voce. In questo modo si può far fronte anche al problema della distorsione.



Cosa fare in caso di una registrazione di bassa qualità

Purtroppo, se la registrazione non raggiunge gli standard indicati la trascrizione ne risentirà pesantemente in qualità. Una volta ottenuto un file audio non è possibile migliorarlo. L'unica soluzione è cercare di ottenere la qualità desiderata nell'atto dell'acquisizione utilizzando i suggerimenti che verranno esposte di seguito.

Riconoscere una buona registrazione

Specifiche tecniche

In questa sezione viene illustrata la lista dei formati e delle varie configurazioni che i file audio devono avere per essere idonei al nostro scopo.

La fase di raccolta dell'audio è la prima fase del processo operativo che porta alla trascrizione da voce a testo. Non esiste un "sistema migliore" per effettuare la raccolta; dipende dal contesto. La raccolta può avvenire mediante vari sistemi e tecniche.

Il sistema FlyScribe può operare su file di tipo WAV, MP3 o OGG, con caratteristiche minimali come di seguito definite:

- **WAV**: 16kHz (sampling rate), 16bit, 1 canale (mono) di tipo RIFF/PCM (lo standard);
- **MP3** o **OGG**: 16kHz (sampling rate), mono o stereo, almeno 64kbps bit-rate.

Le caratteristiche descritte sono "minimali" nel senso che i parametri numerici non possono essere ridotti al di sotto dei valori proposti; questo non vuol dire che l'incremento degli stessi dia però alcun beneficio. Registrare a 44kHz o in stereo non dà alcun beneficio (ed è dannoso, come spiegato sopra, se si mantiene il bitrate invariato). Di fatto, file con caratteristiche diverse vengono, ove possibile, automaticamente convertiti alle caratteristiche di base.

Si noti che il bit-rate non è una misura assoluta della qualità dell'audio: un file registrato a 44kHz e compresso a 64kbps avrà una qualità quasi tre volte peggiore del caso ideale a 16 kHz. Il valore minimo di 64kbps vale perciò unicamente per i file a 16kHz, 16 bit, mono.

Le proprietà di un file audio

In Windows, le proprietà di un file audio possono essere visualizzate selezionando il file audio e, premendo il tasto destro del mouse, accedendo alla voce "Proprietà".



Le caratteristiche tecniche di un file audio possono essere visualizzate in modo differente, a seconda della presenza sul proprio personal computer di programmi aggiuntivi. Negli esempi che seguono il programma che migliora l'interfaccia utente è il prodotto gratuito di conversione di formato dbPowerAmp.

Di seguito vengono riportati tre esempi relativi a due file audio. I primi due sono esempi rilevati in assenza di prodotti esterni.

La figura 1 riporta la finestra delle Proprietà relative ad un file audio WAV. All'interno di questa sono presenti due ulteriori finestre: Generale e Riepilogo. Le specifiche vengono rese disponibili nella finestra Riepilogo nella modalità Avanzata (azionabile mediante il comando presente nella finestra).

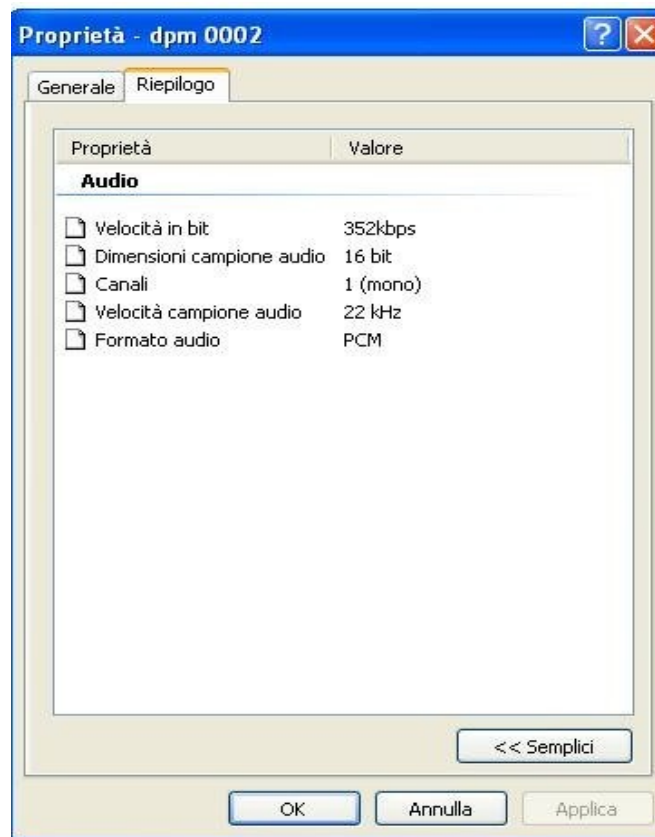


Figura 1: Finestra di Proprietà relativa ad un file audio WAV

Nel caso della figura 2, le proprietà sono relative ad un file audio MP3. Le specifiche vengono sempre rese disponibili nella finestra Riepilogo nella modalità Avanzata come visto nel caso precedente.



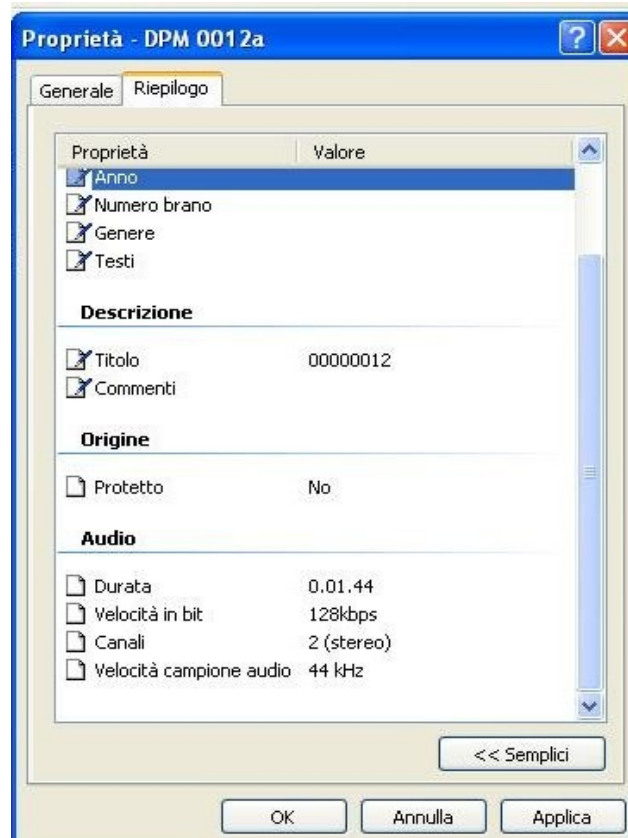


Figura 2: Finestra di Proprietà relativa ad un file audio MP3

La situazione che si può vedere in figura 3 è riscontrabile nel caso venga installato il sopraccitato dbPowerAmp. In questo caso la finestra Proprietà ha più sotto finestre tra cui "Audio Properties" che evidenzia le specifiche del file audio. Grazie all'utilizzo di questo software ci sono informazioni approfondite.



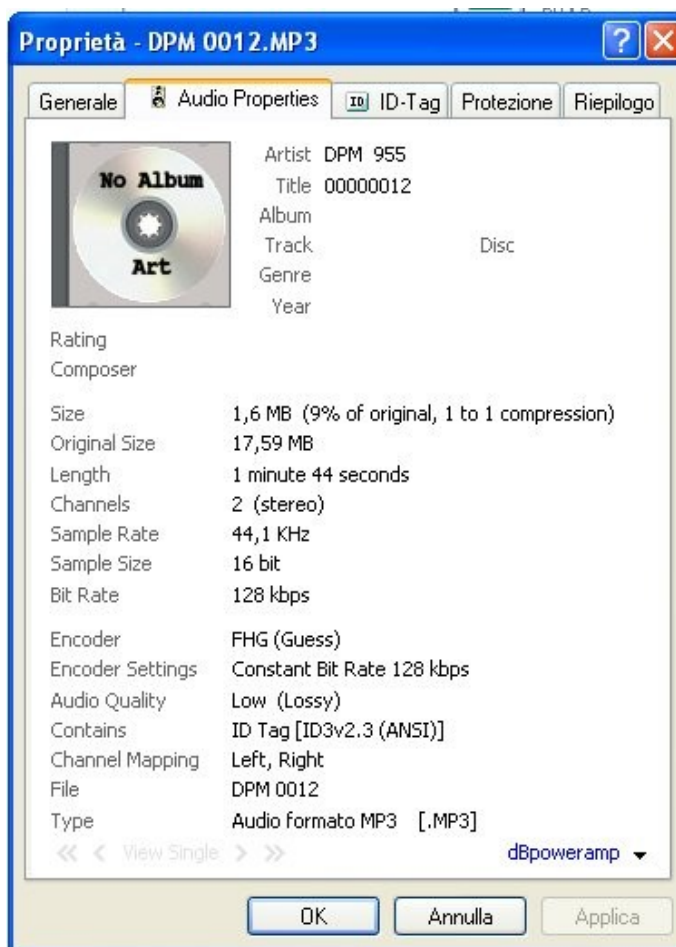


Figura 3: Finestra di Proprietà relativa ad un file audio MP3 con dbPowerAmp

Software utili

Qui presentiamo alcuni programmi molto utili per la registrazione e gestione di file audio. Tenere presente che ci sono innumerevoli programmi sia gratis che a pagamento che possono essere utili. Tuttavia questi strumenti sono di buona qualità e di fruibilità immediata.

Audacity

Audacity (<http://audacity.sourceforge.net>) permette la registrazione, la riproduzione, la modifica e il mixaggio di un file audio. L'interfaccia grafica ne permette un utilizzo semplice e intuitivo. Il programma, rilasciato in forma gratuita sotto la licenza GNU General Public License, funziona su vari sistemi operativi (MS Windows, MacOs,



GNU/Linux).

dBPowerAmp

DbpowerAMP (<http://www.dbpoweramp.com>) è uno strumento indispensabile nel caso fosse necessario convertire dei file audio in formati diversi. Inoltre, questo strumento consente di visualizzare svariati dettagli tra le proprietà di un file audio.

L'ambiente di registrazione

Uno degli elementi fondamentali per ottenere una buona registrazione è l'ambiente ovvero la stanza o il luogo dove viene ripreso l'audio. Bisogna infatti cercare di conoscere l'ambiente così da poterlo gestire in modo adeguato.

Riverbero

Un problema rilevante per le registrazioni è quello del riverbero. Il riverbero è un fenomeno acustico legato alla riflessione del suono da parte di un ostacolo posto davanti alla fonte sonora quali per esempio le pareti della stanza, il soffitto e il pavimento. Per esempio, se in una stanza una sorgente sonora cessa di irradiare, il livello sonoro diminuisce tanto più lentamente quanto minore è l'assorbimento acustico delle pareti.

Il riverbero ha aspetti negativi, come il rischio di mascheramento delle sillabe del parlato o del fraseggio musicale. Per la musica ha aspetti positivi, come il rinforzo dell'intensità della sorgente. Purtroppo tale rinforzo non migliora la qualità del parlato.

Quindi per risolvere questo tipo di problematica è utile ridurre la produzione. Il problema del riverbero può essere attenuato chiudendo le tende (se presenti). In alternativa si tenta di attenuare il problema con il riorientamento dei microfoni. Se, nonostante questi provvedimenti, il problema fosse ancora forte è indicata l'installazione di pannelli fonoassorbenti.

Eco

Si parla di eco quando le singole riflessioni dell'onda sonora sono percepite distintamente dall'ascoltatore. La presenza di un'eco può pregiudicare notevolmente la qualità della trascrizione. È molto difficile attenuare questo problema in quanto direttamente correlato alle dimensioni della stanza. Comunque l'introduzione di pannelli fonoassorbenti può risultare efficace.



I tipi di microfono

Per registrazioni in stanze, sale o aule di assemblea si adoperano generalmente tre tipi di microfono, classificati secondo il tipo di sensibilità direzionale: unidirezionale, superdirezionale e omnidirezionale. Un corretto impiego dei microfoni richiede una conoscenza delle loro caratteristiche elettroacustiche.

Unidirezionale: cardiode

Il tipico esempio di microfono da sala è quello unidirezionale, solitamente di tipo cardioide. La particolare caratteristica del microfono a cardioide (elevata sensibilità dei suoni provenienti dal davanti e dai fianchi contro una brusca attenuazione di quelli provenienti dal retro) è molto vantaggiosa in alcune applicazioni, per esempio, se la registrazione viene effettuata in una stanza.

Un microfono direzionale ha una sensibilità che varia a seconda della direzione di provenienza del suono. Questo tipo di microfono è sensibile in modo significativo solo ai suoni provenienti da una direzione frontale. In tale modo, quando vi è molto rumore, è possibile puntare il microfono in una certa direzione per raccogliere i suoni desiderati ed ignorare i rumori provenienti da altre direzioni. La maggior parte dei microfoni oggi in uso è di questo tipo.

Unidirezionale: supercardioide

Il supercardioide possiede un angolo direzionale molto stretto ed è particolarmente adatto a raccogliere suoni provenienti da una ben precisa direzione. Usando questo microfono, l'operatore può puntarlo verso la sorgente sonora desiderata. Questo microfono è indicato nel caso in cui i parlatori mantengano lo stesso orientamento nei confronti del microfono. Nel caso in cui un parlatore si sposta, anche solo leggermente, il microfono tende a non ricevere il segnale. Inoltre questo tipo di microfono estromette rumore provocato da chiacchiere o brusio in prossimità del parlatore.

Omnidirezionale

L'uso di questo tipo di microfono può essere efficace in sale piccole con pochi parlatori. Il microfono omnidirezionale fornisce lo stesso segnale indipendentemente dalla direzione di provenienza del suono. Questo tipo di microfono è molto adatto per uso comune. È inoltre particolarmente adatto ad essere usato da chi non abbia molta esperienza nell'impiego di microfoni o per raccogliere tutti i suoni di un ambiente durante registrazioni dal vivo. Tuttavia poiché non è possibile puntare il microfono omnidirezionale verso una sorgente sonora, non è adatto a raccogliere suoni particolari in un ambiente con rumore o riverbero.



Le fonte di rumore

Per rumore di fondo si intende il brusio, il rumore elettrico, ed il rumore elettronico presente nella registrazione. Il brusio può essere rimosso facendo rispettare il silenzio nell'aula (anche se in alcuni casi può risultare difficile).

Il rombo di origine elettrica, il soffio dell'amplificatore, lo stridio creato da sistemi digitali costituiscono i rumori elettrici.

Per quanto riguarda il rombo elettrico, non c'è un solo modo per rimuovere questo genere di disturbo. È necessario che l'impianto di amplificazione sia costruito secondo le norme di sicurezza e che le masse siano corrette ed efficaci. Controllare in ogni caso la corretta disposizione dei cavi. Un possibile consiglio è quello di invertire la spina della corrente elettrica. Il problema del soffio dovuto all'amplificatore può essere ignorato, se non troppo evidente. Se il problema però dovesse essere troppo fastidioso, una soluzione è quella di attenuare un po' gli alti.

Lo stridio creato da sistemi digitali è un problema solitamente dovuto alla conversione del file audio. E' consigliato registrare direttamente nel formato definitivo, perchè spesso, durante le conversioni, avviene un fenomeno (aliasing) che crea dei fastidiosi suoni spuri.

Nel capitolo dedicato al riconoscimento di una buona registrazione, saranno trattate le specifiche consigliate per i file audio per la trascrizione automatica eseguita con Audioma.

I problemi legati al parlatore

Ci sono tre problemi principali legati al parlatore: l'emotività, la sovrapposizione delle voci, e l'uso di dialetti.

Emotività del parlato spontaneo

Mentre nel caso ideale un sistema di trascrizione opera su una registrazione di una persona che parla normalmente, nella realtà ci sono tutta una serie di problematiche legate all'emotività del parlatore.

Prima di tutto, ogni forma di emozione può portare ad una istintiva variazione di tono e di volume del parlato di ciascuno di noi.

Inoltre, ci sono dei comportamenti dovuti all'emotività, tali per cui alcuni suoni non sono ripresi correttamente. Questo succede quando il parlatore si muove e non parla direttamente nel microfono.



I problemi che emergono sono dunque:

- **distorsione.** Questo problema si evidenzia quando un parlatore alza il volume della voce oltre ai livelli tarati all'inizio della registrazione;
- **parole troncate e incomprensibili.** Occorre quando un parlatore non parla nel microfono o quando abbassa troppo il volume della propria voce.

Queste alterazioni possono avere impatti negativi sulle prestazioni di riconoscimento automatico. In entrambi i casi è importante impostare i livelli minimali e massimali del volume in entrata, ma per definizione è difficile prevedere o prevenire comportamenti emotivi.

Sovrapposizione di più parlatori

Nel corso di una registrazione è possibile che più parlatori sovrappongano le loro voci. Questo problema può essere risolto utilizzando canali di registrazione separati per ogni parlatore oppure utilizzando mixer automatici in grado di evitare che canali diversi siano amplificati contemporaneamente, assegnando l'amplificazione ad un solo microfono per volta secondo uno schema o una regola.

In questo caso FlyScribe non riesce a distinguere distintamente le parole, quindi molto spesso riconosce parole fuori dal contesto.

Parlato dialettale

Per parlato dialettale si intende l'espressione in termini non censiti nel vocabolario in uso da FlyScribe ed eventualmente con fonemi non presenti nella lingua italiana. Il sistema cercherà in questo caso di rendere la parola acusticamente più "vicina" tra quelle a lui note, che spesso non ha alcuna relazione con la parola veramente pronunciata.

In linea generale, il parlato dialettale non viene gestito. Particolari espressioni usate frequentemente possono essere introdotte nel vocabolario, purché esista una casistica rappresentativa.

Ulteriori informazioni

Per ulteriori informazioni contattare l'assistenza tecnica di PerVoice S.p.A. tramite email: assistenza@pervoice.it o telefono +39 0461 431 355.

