Appendice A: Introduzione al Digital Audio

Digital Audio di Base

Nel mondo digitale tutto è ridotto a uno stato acceso o spento in modo che possa essere memorizzato nella memoria del computer come un singolo bit di informazione: 1 o 0. Cose reali complesse, come immagini e audio, non possono essere rappresentati direttamente in maniera così semplice. L'immagine è raramente composto di punti bianchi e neri e l'audio è raramente esattamente acceso o spento.

Ridurre immagini e audio ad uno stato digitale richiede una conversione analogico- digitale . Invece di utilizzare un solo bit di informazione , molti bit sono utilizzati per memorizzare più precisamente lo stato. Usando 2 bit, per esempio, si hanno quattro stati: 00, 01, 10, 11. Per le immagini potrebbero essere nero (00) , grigio scuro (01) , grigio chiaro (10) e bianco (11). Per l'audio si avrebbero quattro diversi livelli di volume. In genere, vengono usati molti più bit. La maggior parte delle schede video del computer usano da 16 a 32 bit per memorizzare un singolo punto. Le schede audio utilizzano in genere 16 bit per i livelli audio.

Il numero di bit da usare dipende dalla percezione umana e l'allineamento dei bit all'interno del computer . I computer tendono a combinare i bit in gruppi di 8, chiamati byte, in modo da utilizzare 8, 16, 24 o 32 bit che si adatterebbero bene in 1, 2, 3, o 4 byte, rispettivamente. Per le immagini, 16 bit non forniscono abbastanza stati per realizzare transizioni impercettibili da uno stato all'altro, quindi si utilizzano 24 o più bit. Per l'audio 16 bit sono adeguati, che è quello che utilizza un CD ma i sistemi audio che utilizzano 24 bit saranno comuni in futuro.

L'audio digitale consiste di migliaia di numeri, chiamati campioni. Ciascun campione contiene lo stato, o ampiezza (volume), di un suono in un dato istante di tempo. Per le immagini, ogni punto di luce, o di pixel, ha una certa luminosità e posizione e tutti i pixel si combinano per fare una foto (vedi figura sotto). Per l'audio digitale, tutti i campioni si combinano per fare una forma d'onda del suono.

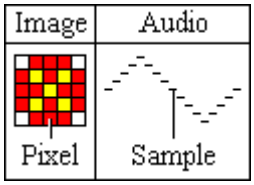

Figura: Esempi

Durante la riproduzione audio, ogni campione specifica la posizione del diffusore in un determinato momento. Un piccolo numero sposta l'altoparlante all'interno e un gran numero sposta l'altoparlante all'esterno. Questo movimento si verifica migliaia di volte al secondo, causando vibrazioni, che si sentono come suono.

Caratteristiche dell' audio digitale

Ci sono diversi caratteristiche che determinano la qualità e la dimensione di un file audio digitale. Sono la frequenza di campionamento, il numero di bit, il numero di canali e il bitrate.

Sampling Rate (frequenza di campionamento)

La frequenza di campionamento è il numero di volte al secondo che il livello di ampiezza (o stato) viene acquisito. Si misura in Hertz (secondi^{-1}, Hz). Un alto valore della frequenza di campionamento determina un suono digitale di alta qualità allo stesso modo in cui l'alta risoluzione video mostra una migliore qualità dell'immagine. I CD, per esempio, usano una frequenza di campionamento di 44100 Hz, mentre i sistemi telefonici usano una frequenza di soli 8000Hz. Se hai mai sentito la musica sul telefono, mentre sei in attesa, avrai notato una grande differenza di qualità rispetto alla musica originale riprodotto su un lettore CD.

Frequenze di campionamento più elevate catturano una gamma più ampia di frequenze e mantengono una forma d'onda liscia. La figura seguente mostra la forma d'onda reale in rosso e la forma d'onda digitale in nero a diverse frequenze di campionamento. Si può vedere che l'aumento del tasso di campionamento rende ristretta ogni fase della forma d'onda digitale. La forma segue più da vicino il mondo reale. In generale, l' altezza di ogni gradino è ridotto e dipende dal numero di bit. In termini semplici, la frequenza di campionamento controlla la larghezza di ciascun gradino.

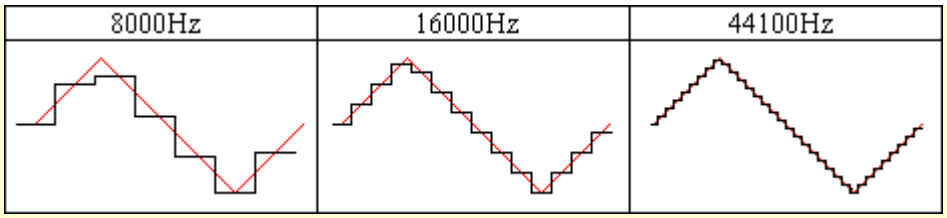

Figure: Sampling Rate

Il tasso da utilizzare dipende dal tipo di suono e la quantità di spazio disponibile. Tassi più elevati consumano molto spazio. Nell'esempio precedente il CD richiede più di 5 volte la quantità di memoria rispetto al sistema telefonico per lo stesso suono digitale.
Alcuni tipi di suoni possono essere registrati a frequenze inferiori, senza perdita di qualità. Alcune frequenze standard sono elencate nella tabella.

Tabella: Campionamenti Standard

Attributi	**Qualità ed utilizzo**	**Memorizzazione (16 bit/mono, MB/minuto)**
8000Hz	Low quality. Used for telephone systems. Good for speech. Not recommended for music.	0.960
11025Hz	Fair quality. Good for speech and AM radio recordings.	1.323
22050Hz	Medium quality. Good for TV and FM radio quality music.	2.646
44100Hz	High quality. Used for audio CDs.	5.292
48000Hz	High quality.	5.760

	Used for digital audio tapes (DAT).	
96000Hz	Very high quality. Used for DVD audio.	11.520

Numero di bit

Come spiegato nella sezione Digital Audio Basics, il numero di bit determina quanto accuratamente l'ampiezza della forma d'onda viene catturata. La figura seguente mostra una vera e propria forma d'onda in rosso e le corrispondenti forme d'onda digitali con campioni a 2 bit e campioni a 3 bit.

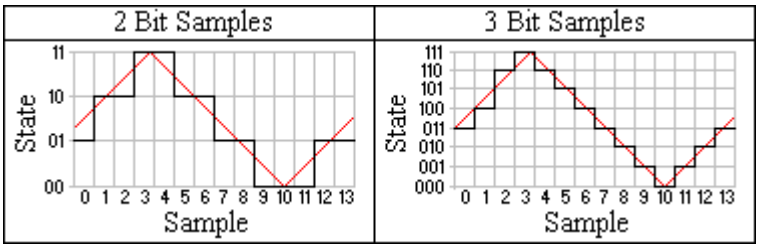

Figura: Bits

Si può vedere che l'aggiunta di un singolo bit migliora notevolmente il modo in cui la forma d'onda digitale è conforme alla vera forma d'onda. La forma d'onda a 2 bit si presenta come una rozza approssimazione con grandi gradini. Diverse ampiezze sono approssimate allo stesso valore, come i campioni da 9 a 11 . Questa è una fonte di rumore di quantizzazione, spiegato più avanti.
Nella forma d'onda a 3 bit, nessuna ampiezza è arrotondata allo stesso valore. Ogni passo è metà dell'altezza della forma d'onda a 2 bit, ma non è ancora perfetto. Dal campione 1 a campione 2 , c'è un salto nella forma d'onda, che provoca anche rumore di quantizzazione in misura molto minore. Noterete che i campioni 0 e 1 sono sotto la vera forma d'onda e campioni 2 e 3 sono sopra la forma d'onda. Ciò si verifica perché non ci sono di mezzo gli stati per memorizzare con precisione quei livelli di ampiezza, quindi la forma d'onda digitale finisce a cavallo tra quella reale. Pertanto sono necessari ulteriori stati e altri bit.
Sono comuni campionamenti a 8 bit e 16 bit. In un campione di 8 bit, ci sono 256 stati o livelli di ampiezza diverse. Campioni a 16 bit hanno 65.536 livelli. Questo fa una grande differenza in termini di qualità del suono. L'audio memorizzato come campioni di 8 bit avrà spesso molto più rumore di quantizzazione.

Signed and Unsigned (valori con segno e senza segno)
I campioni possono essere conservati come bit in due modi diversi. Un primo modo è quello di considerare tutti gli stati positivi, senza valori negativi. Come mostrato nella figura precedente, gli stati 00, 01, 10, e 11 sono uguali ai numeri positivi 0, 1, 2 e 3. Questo elimina la necessità di un segno negativo. Questi campioni sono chiamati unsigned. Per i campioni di 8 bit, gli stati varierebbero da 0 a 255.
L'altro modo è quello di utilizzare una forma nota come complemento a due, che consente valori sia positivi che negativi. Questi campioni sono chiamati senza segno. Poiché le forme d'onda del mondo reale tendono a fluttuare attraverso una gamma di valori positivi e negativi, campioni con segno sono preferiti. Per i campioni a 16 bit, gli stati variano da -32.678 a +32.767.

Big and Little Endian

Quando un campione viene memorizzato utilizzando più di 8 bit, è necessario più di un byte. Il termine endian è usato per descrivere il modo in byte ordinati nella memoria del computer. Si specifica il significato del primo byte del gruppo. Un campione di 16 bit, per esempio, richiede esattamente due byte, byte A e byte B. Essi possono essere memorizzati come A prima, poi B o B prima, quindi A. Generalmente un PC li memorizza in un modo e un Mac salverà nell'altro modo a causa di differenze nel processore interno di tali sistemi.

Nel formato **Big Endian** è memorizzato per primo il byte più significativo, che lo rende simile al modo in cui leggiamo i numeri. Nel numero 47, il 4 è il primo ed è il più significativo e il 7 è l'ultimo ed è meno significativo. Questo ordinamento viene utilizzato su sistemi Mac.

Nel formato **Little Endian** il byte meno significativo è memorizzato prima, permettendo alcune ottimizzazioni nella trasformazione. Questo ordinamento viene utilizzato su sistemi PC Intel.

Canali

L'audio digitale può avere uno o più canali. Canale audio singolo, indicato come un audio monofonico (o mono), contiene le informazioni per un solo altoparlante ed è simile alla radio AM. Due canali audio o audio stereo, contiene i dati per due altoparlanti o due orecchie, proprio come FM stereo. Suoni stereo possono aggiungere profondità, ma richiedono il doppio del tempo di lavorazione e memorizzazione come suoni mono. La maggior parte delle sale cinematografiche utilizzano sistemi audio avanzati con 4 o più canali, che sono in grado di rendere i suoni provenienti da specifiche direzioni. GoldWave supporta attualmente solo l'audio mono e stereo.

Limiti del Digital Audio

Poiché l’audio digitale è limitata dalla frequenza di campionamento, il numero di bit e il numero di canali, una forma d'onda digitale non può essere mai una replica esatta della forma d'onda reale. Queste limitazioni possono portare a una serie di problemi, come aliasing, clipping e rumore di quantizzazione.

Aliasing

L’aliasing si verifica quando la frequenza di campionamento non è sufficientemente elevata per catturare correttamente la forma dell'onda sonora. Il suono registrato avrà toni mancanti (Figura: Aliasing, sopra) o nuovi toni che non sono mai esistiti nel suono originale (Figura: Aliasing, basso). Questo problema può essere eliminato utilizzando frequenze di campionamento più elevate o utilizzando filtri anti-aliasing.

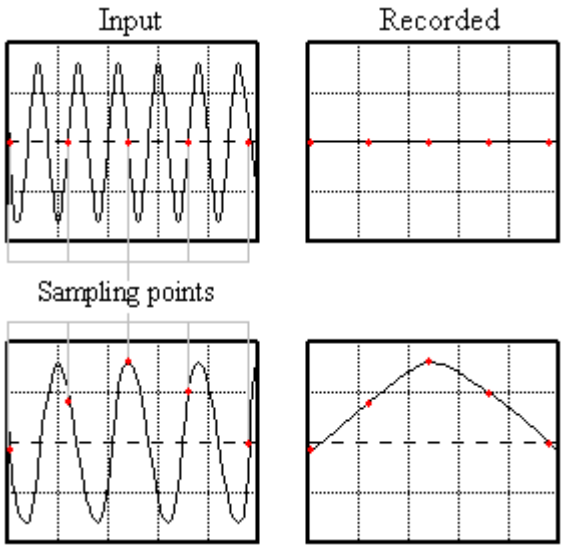

Figure: Aliasing

Frequenze di campionamento più elevate aumentano il numero di punti di campionamento. Per vedere come funziona, prova ad aggiungere tre punti tra ogni punto di campionamento nella figura e ridisegna il grafico. L'audio registrato viene più simile all'ingresso.

I filtri anti-aliasing rimuovere tutti i toni che non possono essere campionati correttamente. Impediscono che i toni molto acuti vengano alias di tono basso. Molte schede audio includono filtri anti-aliasing nell'hardware.

Clipping

Errori di clipping si verificano quando il livello è troppo elevato per essere memorizzati nei bit disponibili. Ad esempio, se il livello massimo per un campione di 16 bit è 32767 e il livello effettivo è 40000, allora deve essere tagliato a 32767 (vedi figura). Questo genera distorsione. Per eliminare il taglio, regolare il volume di registrazione prima di registrare. Utilizzando la finestra di controllo "Monitor input on visuals feature", è possibile regolare il volume ad un livello adeguato. Il volume è abbastanza basso quando il misuratore VU visivo non raggiunge la parte superiore della regione rossa.

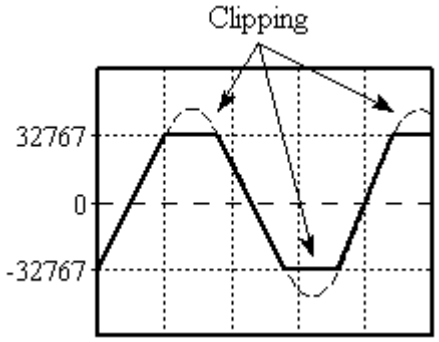

Figure: Clipping

Clipping può verificarsi quando si elaborano effetti, come l'aumento del volume con l'effetto Variazione Volume. GoldWave non taglia l'audio internamente, ma deve essere tagliato quando si invia l'audio alla scheda audio o quando si salva un file. Si consiglia di utilizzare l'effetto

Massimizzare Volume con una impostazione a 0 dB prima di salvare per garantire che nessun clipping si verifichi nel file.

Quantization

Il mondo reale ha un numero infinito di stati, ma il mondo digitale ha un numero molto limitato di stati. La quantizzazione è un processo di assegnazione di una ampiezza di questa quantità limitata di stati. Il rumore di quantizzazione si verifica quando una ampiezza è arrotondato allo stato più vicino per il dato numero di bit. Questo è illustrato nella sezione bits. L'uso di 2 bit provoca grandi errori di arrotondamento e segue solo approssimativamente la vera forma d'onda. L'utilizzo di 3 bit ha errori di arrotondamento molto più piccoli ma è ancora molto lontana dalla vera forma d'onda. L'utilizzo di 8 bit fornisce una forma d'onda molto più agevole, ma gli errori di arrotondamento possono ancora essere sentiti. L'utilizzo di 16 bit dà errori così piccoli che sono quasi impercettibili.

Noise

Per ridurre al minimo i rumori interni ed esterni, assicurarsi che la scheda audio sia installata il più lontano possibile dalla scheda grafica. Se l'hardware audio è integrato nella scheda madre del computer, considerare l'acquisto e l'installazione di una scheda audio separata invece. Hardware integrata tende ad avere un livello di rumore molto più alto.

Tenere tutti i microfoni e cavi audio lontano dal monitor, case del computer, o di altre fonti di rumore elettrico. Utilizzare cavi schermati e assicurarsi che tutto abbia una connessione di terra comune. Collegare le apparecchiature e dispositivi nella stessa ciabatta in modo che condividano la stessa terra. Apparecchiature senza una spina con messa a terra (spina a tre punte) può richiedere la messa a terra speciale. A volte ruotando la spina aiuta a ridurre un ronzio o brusio.

La configurazione del sistema può anche compromettere la qualità audio. A causa di problemi architettonici con PC ed eccessiva memoria virtuale di scambio con Windows, è possibile notare un gap occasionale durante la registrazione. Il riavvio di Windows, l'aggiornamento del driver audio, chiudendo tutti gli altri programmi, o l'installazione di più RAM aiuta a minimizzare questo problema.

Volume Scales

I livelli di forma d'onda o stati possono essere interpretate in un numero di modi diversi. La lettura degli stati nella loro forma binaria è impraticabile, così gli stati vengono mappati a differenti scale che sono più leggibili. Questi includono una semplice scala di ampiezza, una scala percentuale, e una scala in decibel. Come la figura mostra, un livello di ampiezza 1.0 è la stessa di un livello 100% e un livello 0dB.

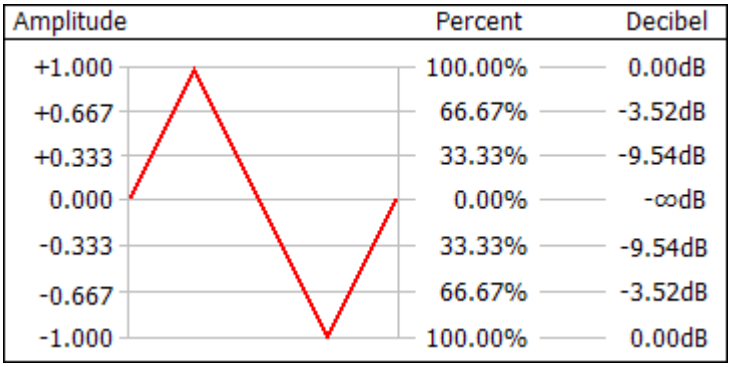

Figure: Volume Scales

Amplitude (y)

La scala di ampiezza associa semplicemente gli stati ad un intervallo lineare da -1,0 a +1,0, con silenzio pari a zero. L'ampiezza tipicamente è data come un valore positivo quando usato in effetti.

Percent (%)

La scala percentuale è del 100% al massimo e 0% in silenzio. È essenzialmente uguale al valore assoluto della scala di ampiezza convertito in una percentuale moltiplicando per cento e aggiungendo un segno di percentuale. Talvolta una percentuale negativa viene utilizzato, come nell'effetto Flanger, per invertire la forma d'onda.

Decibel (dB)

La scala in decibel è insolita in quanto è 0dB al massimo assoluto e all'infinito negativo per il silenzio. Si tratta di una scala logaritmica, che è più vicino al modo uditivo umano che percepisce i livelli sonori. Noterete dalla figura sopra che non ci sono livelli positivi. Livelli di sotto del massimo sono negativi. Solo i valori superiori al massimo sono positivi (non mostrato) e tali valori possono causare clipping. Quando si cambia il volume, valori positivi aumentano il livello e valori negativi riducono il livello.

Usare l'equazione seguente per convertire un livello di decibel ad un livello percentuale.

$$percentage = 10^{\frac{gain(dB)}{20}} \times 100\%$$

Relative vs. Absolute

Regolando il volume in effetti, il valore può essere interpretato come un livello relativo o livello assoluto. Quando si cambia il volume, di solito è relativo. Quando si specifica una soglia, di solito è assoluto. Cambiamenti relativi sono cumulativi. Quindi, se si applica una variazione di volume con 0,5 ampiezza (50 % o - 6.02dB), l' ampiezza diminuisce a metà del livello attuale. Se si applica ancora una volta che il cambiamento, allora scende a un quarto del suo livello originale. In altre

parole, la variazione è relativo al livello attuale. Data un ampiezza originale di A, il primo cambiamento produce un risultato di A x 0,5 e il campione viene sostituito da tale valore. Il secondo cambiamento assume tale valore e lo moltiplica per 0,5 di nuovo, in modo da ottenere il risultato finale di (A x 0,5) x 0,5 o A x 0,25. La maggior parte delle impostazioni di effetti in GoldWave sono cumulativi. Per i cambiamenti relativi, utilizzando 1,0 , 100 % o 0 dB non altera il suono.

Livelli assoluti vengono utilizzati in luoghi in cui è richiesta una soglia, come in Riduzione Silenzio, o in rari casi erano il livello assoluto è impostato, ad esempio massimizzare il volume. Variazioni assolute non sono cumulabili. Se si ingrandisce il suono con 0,5 ampiezza (50% o -6.02dB), allora questo è ciò che sarà il livello di picco, non importa quante volte viene usato l'effetto. Per i cambiamenti assoluti, l'uso di 1,0 , 100% o 0 dB può modificare il suono se al momento non è a quel livello.

Frequency and Pitch

Finestre di suono in GoldWave mostrano suoni come forma d'onda di ampiezza su un asse temporale. Tuttavia, il suono può essere visto in un modo completamente diverso, esaminando la sua frequenza e il contenuto pitch o spettro di frequenza. I suoni sono suddivisi in una combinazione di toni (sinusoidali) fondamentali semplici, ciascuno con una frequenza diversa. Questo è utile per esaminare livelli bassi e acuti o per isolare e studiare certi suoni.

Frequency Ranges

La media dell'udito umano si estende su una gamma di frequenza da circa 20Hz a circa 15000Hz. La figura sottostante mostra alcuni suoni comuni e la gamma di frequenze che coprono.

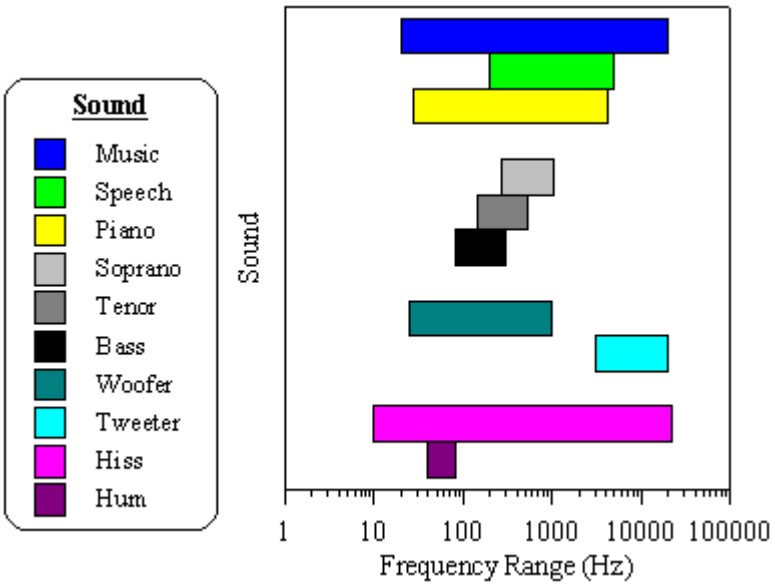

Figure: Frequency Range of Sounds

Molti si chiedono perché è difficile da rimuovere la voce dalla musica. Dalla figura, vedrete c'è una grande sovrapposizione nella gamma di frequenza di parola e musica. Rimozione delle voci potrebbe anche rimuovere una parte significativa della musica. Un problema simile si verifica quando si rimuove il rumore sibilo, dal momento che spesso copre l'intero spettro.

La maggior parte dei sistemi stereo di base hanno controlli dei bassi e degli alti, che offrono un controllo limitato su uno spettro di frequenza. I Bassi applica ai suoni la bassa frequenza, come i tamburi, violoncelli, note basse del pianoforte o un ronzio. Treble applica ai suoni l' alta frequenza, come uno scontro di cembali, un tweet di un piccolo uccello, note alte di un pianoforte o un rumore sibilante. Gli impianti stereo più costosi hanno Equalizzatori Grafici, che forniscono un migliore controllo su uno spettro di frequenza. Invece di controllare solo due bande (bassi e alti), è possibile controllare molte bande. GoldWave fornisce ancora più controllo sulla spettri di frequenza con effetti filtro come Parametric EQ, basso / passa-alto, passa-banda / stop, equalizzatore, riduzione del rumore, e Spectrum Filter.

Frequency Range and Sampling Rate

La gamma di frequenza di un suono digitale è limitata dalla sua frequenza di campionamento. In altre parole, un suono campionato a 8000Hz non può registrare frequenze superiori a 8000Hz. Infatti, il suono non può anche avere frequenze superiori 4000Hz. Secondo il teorema del campionamento, la frequenza massima è superiore alla metà della frequenza di campionamento. Le frequenze più alte conterranno alias, a quelle più basse, provocando rumore se non vengono utilizzati filtri appropriati.

CD audio è progettato per coprire l'intera gamma di udito umano, che ha un massimo al di sotto di 22kHz. Per registrare correttamente questo intervallo, il teorema del campionamento afferma che deve essere utilizzato una frequenza di campionamento di almeno il doppio della massima, così è richiesta una frequenza di almeno 44KHz. La frequenza effettiva è 44100Hz per i lettori CD standard.

Frequency Spectrum Graphs

Molte delle immagini di controllo di GoldWave convertono i suoni in una gamma di bande di frequenza utilizzando un algoritmo radice-2 trasformata veloce di Fourier (FFT). Quando i risultati sono disegnati con colori, il grafico è indicato come uno spettrogramma. Quando i risultati sono disegnate con linee, è spesso indicato come uno spettro di frequenza o analisi di frequenza.
Grafici di analisi di frequenza vengono visualizzati nella riduzione del rumore, Spectrum Filter e effetti filtro Parametric EQ. Questi consentono di individuare le frequenze che si desidera rimuovere o migliorare.
GoldWave applica una funzione di windowing ai dati prima di eseguire la FFT (vedere Controllo sezione Proprietà di Visual). Questo riduce gli errori di "discontinuità" che si verificano quando divide i dati in piccoli pezzi. Una finestra Kaiser viene utilizzato per impostazione predefinita.
Per rendere lo spettro più realistico rispetto all'udito umano, le grandezze sono in scala logaritmica. Ciò significa che se una frequenza "suona" due volte più forte dell'altro, è rappresentata graficamente con due volte l'altezza (o il colore corrispondente per lo spettrogramma).

Compressione dei File

File audio non compressi tendono ad essere grandi. L'audio di qualità CD richiede dieci megabyte al minuto. Questo non è un problema con grandi dischi rigidi dei computer oggi disponibili, ma è un problema se si desidera salvare molte canzoni su un lettore portatile o se si desidera trasferire file su Internet. A differenza di molti dati informatici, i dati audio non si comprimono molto bene usando i metodi di compressione tipici come quelli ottenuti con programmi tipo PKZIP o WinZip. Questi metodi conservano i dati esattamente quindi non c'è perdita di qualità. Tale compressione si chiama compressione lossless.

Per rendere i file audio più piccoli, si devono utilizzare algoritmi complessi. La maggior parte di questi algoritmi sacrificano un po' di qualità in modo che quando i dati vengono decompressi, non si ottiene esattamente la stessa qualità che aveva in origine. Questo tipo di compressione è noto come compressione lossy. Idealmente la qualità che si perde non è percepibile e non si nota la differenza. Il metodo più comune di compressione audio lossy è MPEG Layer- 3, meglio noto come MP3 . È in grado di ottenere audio di qualità vicino CD in meno di un decimo della dimensione, che è circa un megabyte al minuto. Algoritmi più recenti, quali AAC, Ogg Vorbis e Windows Media Audio ottengono anche una migliore qualità in un formato più piccolo. Software e hardware che comprimono l'audio utilizzano algoritmi complessi che sono indicati come codec (coder/decoder). La compressione è come la codifica e la decompressione come la decodifica. Per controllare il livello di compressione in GoldWave, usa File | Salva con nome e selezionare un tipo e/o gli attributi di file diverso. Più basso è il bitrate (il numero Kbps, vedi sotto), più piccolo sarà il file, di solito a qualità ridotta.

⚠ Durante l'apertura, la modifica e il salvataggio di un file più volte, è meglio usare un formato di file lossless. Ogni volta che un file compresso con perdita di dati è riaperto e salvato, qualche qualità è persa. Per l'editing a lungo termine, utilizzare il tipo di onda con "PCM a 16 bit" attributi di una delle lossless formati compressi come FLAC o Windows Media Audio con attributi "lossless".

Bitrate

Molti formati audio compressi misurano la dimensione come bitrate. Il bitrate è il numero di bit per secondo (bps) necessari per memorizzare l'audio. Solitamente il numero è dato in kilobit o mille bit. Dividere il numero per 8 per determinare il numero di kilobyte necessari per secondo.
La velocità di connessione a internet (larghezza di banda) è spesso misurata in bitrate puro. Un modem a 56K è in grado di ricevere 56 kilobit al secondo. Se vuoi che la gente opti per lo streaming audio MP3 tramite un modem, è necessario comprimere il file con un bitrate massimo di 56kbps. Grazie al protocollo overhead di connessione e Internet, si dovrebbe utilizzare un tasso inferiore per assicurare l'audio che può essere scaricato abbastanza velocemente. Per DSL e connessioni Internet via cavo, si può utilizzare il tasso di 128kbps come l'MP3 standard.
I file audio possono contenere una vasta gamma di suoni, da scontri dei piatti rumorosi al silenzio. Gli algoritmi di solito ottengono compressioni migliori sul silenzio o semplici sezioni audio che sul complesso, audio rumoroso.

Constant Bitrate

Quando si utilizza il bitrate costante, ogni sezione di audio comprime esattamente la stessa dimensione, indipendentemente dal contenuto. Se l'audio contiene silenzio, allora i dati possono essere imbottiti per riempire il bitrate richiesto. Se l'audio contiene musica complessa, quindi la qualità può essere diminuita fino a quando non si inserisce all'interno del bitrate.
Bitrate costante è utile per i sistemi di trasmissione in cui è fissata la velocità di trasmissione.
E' anche facile cercare le posizioni arbitrarie all'interno del flusso audio o file.

Variable Bitrate

Compressione a bitrate variabile utilizza la più piccola dimensione possibile per ogni sezione di audio. Se l'audio contiene silenzio, poi il bitrate sarà molto basso. Se l'audio contiene musica complessa, il bitrate sarà al suo massimo.
Bitrate variabile offre la migliore compressione e qualità. Tuttavia, rende difficile cercare all'interno del flusso o il file poiché non vi è alcuna relazione diretta tra tempo e dimensione.