2. 音樂情緒辨識

2.1 系統架構

我們的系統分為音樂系統和歌詞系統兩部份，音樂系統如圖一所示，將資料分為訓練和測試兩部分，分別抽出音樂特徵，經過訓練，再將測試資料透過我們已經訓練好的模型進行分類。而我們所使用的分類器會在 2.1.3 節中詳述。

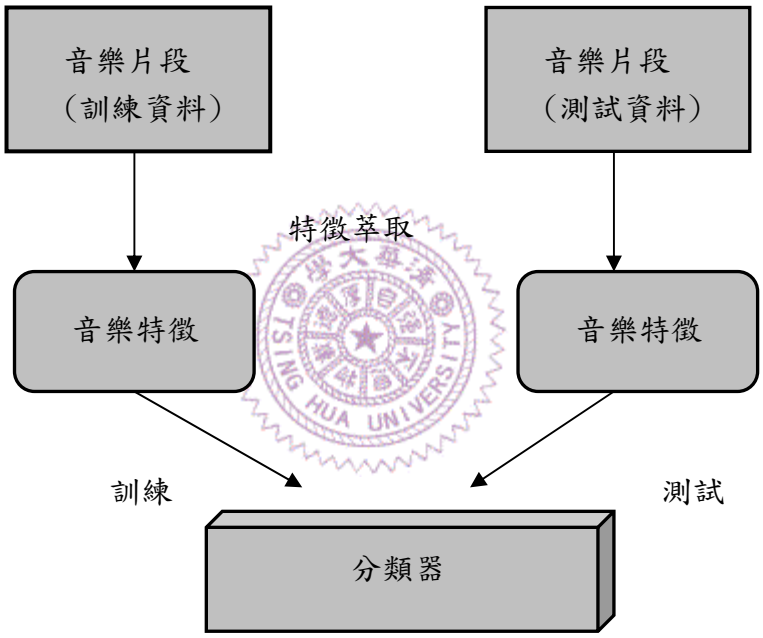

圖 1：音樂內容系統架構

歌詞系統如圖二所示，將資料分為訓練和測試兩部分，透過斷詞[1]，再經人工選出有意義的情緒詞，並建立同義詞表，存入資料庫。而測試資料便會對我們的情緒資料庫搜尋出符合的字串，透過我們的計算方法，排名出分數最高的類別。計算方法會在 2.1.4 節詳述。

[1] 把輸入的字串分隔成詞串

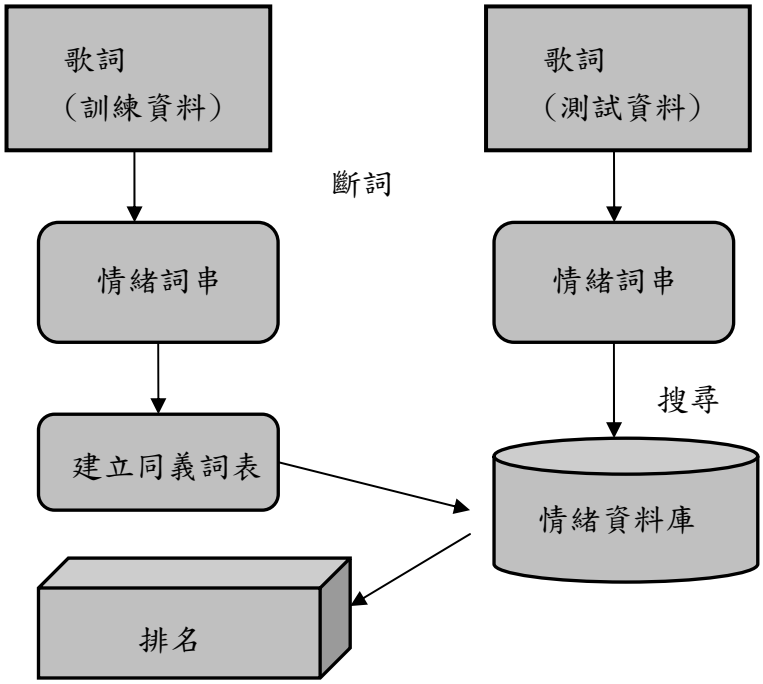

圖 2：歌詞系統架構

2.1.1 歌曲資料庫

本文使用的音樂資料庫是以MIDI（Musical Instrument Digital Interface）[2]轉換出的資訊。由於我們只針對樂曲的旋律內容做分析，所以我們從所有的 MIDI 檔案中萃取和旋律相關(例如:音高、力度、音長等)的資訊。

2.1.2 歌詞資料庫

本文使用的歌詞資料庫是以 KAR [4]檔案（內嵌歌詞的標準 MIDI 檔案）轉換出的資訊。我們從所有的 KAR 檔案中萃取音樂和歌詞兩部份，並將歌詞先經斷詞系統，所謂的斷詞系統[5]是指把輸入的字串分隔成詞串，詞串可以為單字詞或多字詞，例如:輸入詞串為『我到清華大學聽演講』，必須產生正確的字串為「[我][到][清華大學][聽][演講]」，而我們將斷詞後的結果，依照詞串的長度

[2]音樂設備的數位化連接介面

分為二字詞、三字詞、四字詞建立情緒資料表，並統計該情緒字串的詞頻[3]且寫入資料表，再根據情緒詞串以人工建立同義詞詞表。由於歌詞和音樂所要表達的情緒可能有所不同，因此在流行歌曲部份，我們分為 1.只聽音樂 2.只看歌詞 3.聽音樂加歌詞分別標記三組答案。此外，為了驗證模糊理論，我們也針對流行音樂部分，對快樂和焦慮兩類情緒分別標記一組 0~1 之間的答案。

2.1.3 分類器

本論文使用下列四種分類器進行分類，其簡述如下:

2.1.3.1 KNNR

最近鄰居法則(Nearest Neighbor Decision Rule)是指擁有相似特徵的資料，在以其特徵形成的空間中會聚集在一起。也就是說如果把要分類的特徵以高度空間來表示，則屬於同一類的點應該會距離比較近。因此對於一筆未知類別的資料，欲得知其所屬的類別，我們會先將特徵取出，再計算其和訓練資料特徵的距離，我們會判定資料的類別和最接近的點的類別是一樣的。當資料的雜訊較大時，只使用最接近的資料點來判斷可能會失之武斷，因此我們通常會使用 KNNR 進行分類。所謂的 KNNR (K Nearest Neighbor Rule) 是指先求取最接近的 K 個資料點，再根據對應的 M 個類別資訊來決定最後的類別

2.1.3.2 GMM

如果將每一筆資料視為在高維空間中的一點，而這些同一類別的資料點都是由一個高維高斯機率密度函數所產生，就可以用最佳參數估計法 (Maximum Likelihood Estimate) 來求出這個高斯密度函數的最佳參數值。但若資料的分布不是橢球狀，便無法用單一的高斯模型來模擬資料的分布，此時便要用數個高斯模型的加權平均來表示。這種方式就稱為高斯混合模型(Gaussian Mixture

[3] 詞串出現的次數

Model)。因此進行分類時，當高斯模型的加權平均越高，則其屬於該類別的可能性越高，我們會從訓練資料中對於每一個類別的資料，訓練一個 GMM，在測試時 ，將某一筆資料送到每一個類別的 GMM，機率最大者，即代表此資料屬於此類別的可能性較大。

2.1.3.3 SVM

支向機(Support Vector Machine)通常用來處理兩類別的問題，亦可處理多類別的問題。先將訓練資料以+1 或是-1 加以標註，以數學式表示為

$$\{x_i, y_i\}, i = 1 \ldots l, y_i \in \{1, -1\}, x_i \in R^d \quad (1)$$

假設有一個超平面可以將+1 及-1 的資料加以區分，則此超平面就可稱為區分平面(Separating Hyperplans)，若在此超平面上的 x 必須滿足：

$$w \cdot x + b = 0 \quad (2)$$

w 為超平面的法向量，其可用圖 3 表示。而支向機的目標是要在高維度的特徵空間，找出一個具有最大邊界(margin)[4][6]的區分平面來區分兩類資料。

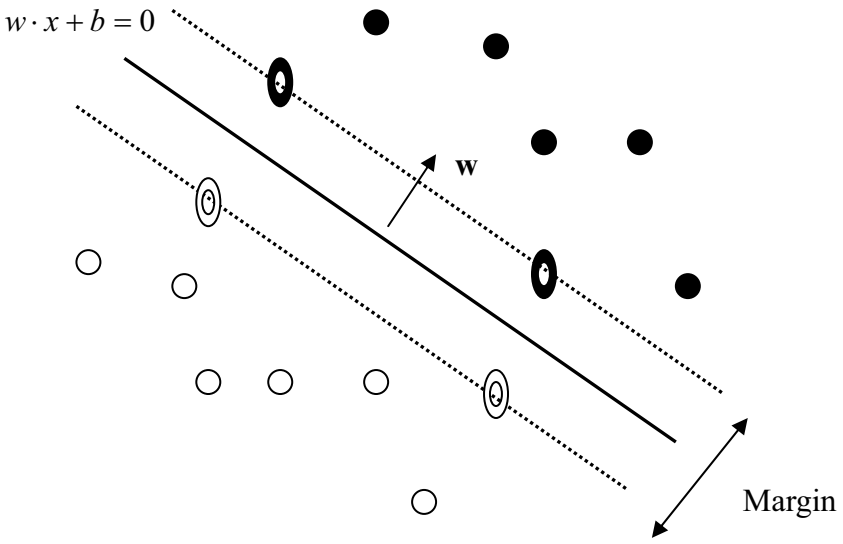

圖 3：支向機

[4] 訓練資料跟區分平面的最短距離

2.1.3.4 Fuzzy KNNR

模糊理論是一門用以將模糊概念量化的學問，起源於1965年Zadeh所發表的「模糊集合」[7]。模糊理論以模糊集合為基礎，以研究不確定事物為目標，接受模糊現象存在的事實，根據不清晰訊息，透過近似推理（ Approximation Reasoning）[8]過程而得到正確結果。

傳統的明確集合（Crisp Set)的特徵函數（Characteristic Function)採用二分法，如表一所示；而模糊理論的精神則將其擴展到 0~1 之間的值，稱為歸屬函數（Membership Function)，如表二所示。若歌曲屬於某類別的程度越大，其歸屬程度越接近 1。

	快樂類別	焦慮類別
歌曲 1	0	1
歌曲 2	1	0

表格 1：傳統明顯集合的特徵函數值

	快樂類別	焦慮類別
歌曲 1	0.2	0.8
歌曲 2	0.9	0.1

表格 2：模糊理論的歸屬函數值

由於 KNNR 會將屬於同一類別的每筆資料都視為同等重要，在分類含有情感的資料時，便會過於主觀。因此本論文採用 Keller 在 1985 年所提出的模糊最近鄰居法則（Fuzzy K-Nearest Neighbor Algorithm)[9]，其以 KNNR 理論為基礎，但測試資料會依其與所取最接近 k 個訓練資料點和這些訓練資料之歸屬程度的距離來決定，屬於同一類別的資料會依距離的遠近而有不同的歸屬程度。其歸屬程度介於 0~1 之間。

2.1.4 歌詞計算方法

本論文將歌詞透過斷詞處理之後，和我們所建立好的情緒資料庫做搜尋，若搜尋到，會分別採用下列三種計算方法計算，並在實驗部份比較其效能。

2.1.4.1 計算方法一

我們假設每一個情緒詞串不是同等重要，因此將測試資料經過斷詞處理拆解成二字詞、三字詞、四字詞，以這些詞組對資料庫做搜尋，會得到一個二~四字詞對應類別的詞頻表，將詞頻乘上該詞所對應的權重，分別加總該歌詞所搜尋到情緒詞的分數，類別分數最高者，則為該類。在本論文中，二字詞權重為 2;三字詞權重為 3;四字詞權重為 4。其計算方法如下:

1. 計算對資料庫搜尋所得到的情緒詞分數

$$score = f_{d,t} * W_i \tag{3}$$

$f_{d,t}$ = 詞 T 在類別 D 出現的次數

w_i =詞 T 所對應的權重

2. 將所計算出的情緒詞分數相加

$$score(d) = \sum_{1}^{j} score_j \tag{4}$$

J 為在 D 類別搜尋到的情緒詞總數

3.找出類別分數的最大值，則為該類。

2.1.4.2 計算方法二

和計算方法一相似，但假設每個情緒詞都是同等重要，因此不乘上詞頻。將測試資料經過斷詞處理拆解成二字詞、三字詞、四字詞，以這些詞組對資料庫做搜尋，若搜尋成功，則視為 1，再乘上一個權重。在本論文中，二字詞權重

為 2；三字詞權重為 3；四字詞權重為 4。其計算方法如下：

1. 計算對資料庫搜尋所得到的情緒詞分數

$$score = 1 * W_i \tag{5}$$

w_i =詞 T 所對應的權重

2. 將所計算出的情緒詞分數相加

$$score(d) = \sum_{1}^{j} score_j \tag{6}$$

J 為在 D 類別搜尋到的情緒詞總數

3.找出類別分數的最大值，則為該類。

2.1.4.3 計算方法三

計算方法三是利用 Managing Gigabytes 一書中所提到的 TF*IDF rule (Term Frequency, Inverse Document Frequency)[10]，其計算方法如下

1. 計算對資料庫搜尋所得到的情緒詞分數

$$score = f_{d,t} \cdot \log \frac{N}{f_t} \tag{7}$$

$F_{D,T}$= 詞 T 在類別 D 中出現的次數

N = 總類別數

F_T = 出現過詞 T 的類別數

2. 將所計算出的情緒詞分數相加

$$score(d) = \sum_{1}^{j} score_j \tag{8}$$

J 為在 D 類別搜尋到的情緒詞總數

3.找出類別分數的最大值，則為該類

這樣做法的意義在於，通常衡量重要性，皆是以該詞在類別內出現的次數做為決定性因子(TF 的含義)，但若該詞同時出現在多個類別中，相對而言該詞比

出現在少數的類別的詞彙較不具價值(IDF 的含義)。

2.2 情緒分類模型

要分類情緒之前需要先建立好情緒分類的模型。本篇論文是採用 1990 年 Thayer[11]所提出的二維情緒模型為基礎，如圖 4 所示。

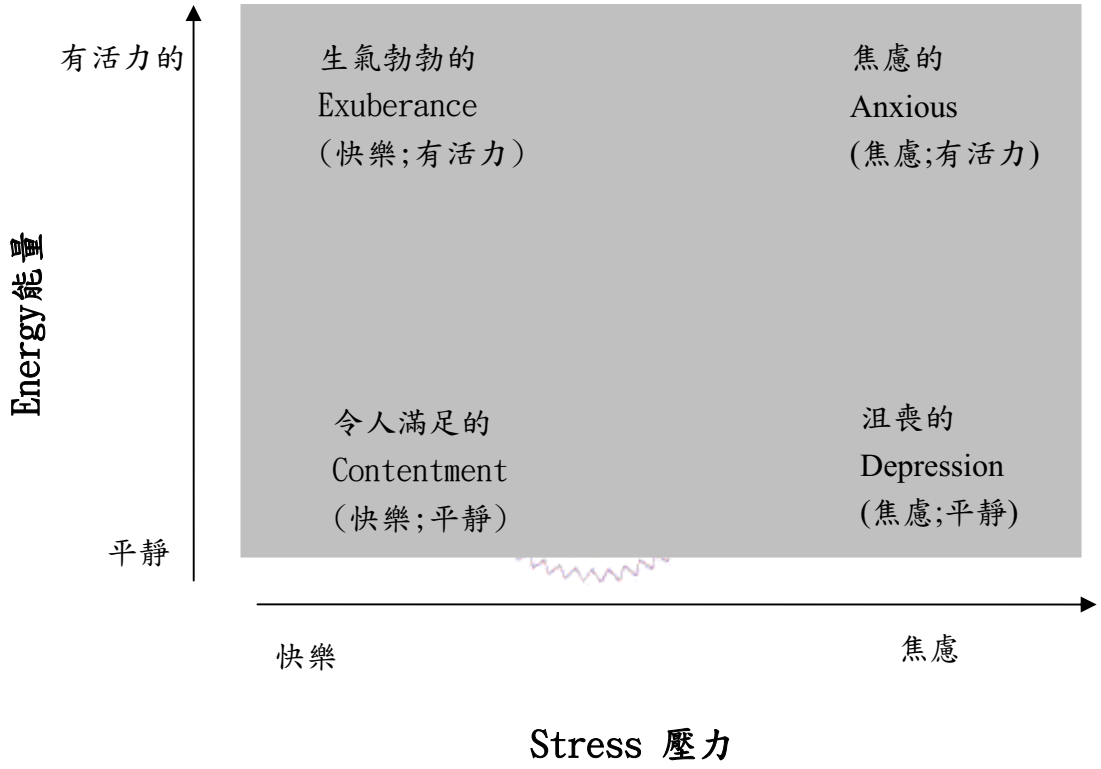

圖 4：Tayer 情緒分類模型

Thayer 的情緒模型主要著重兩個特徵：能量（Energy）和壓力（Stress），因此在本篇論文的情緒模型中我們將能量用兩個情緒詞(「有活力(Energetic)」、「平靜(Calm)」)來表示；壓力用兩個情緒詞(「快樂(Happy)」、「焦慮(Anxious)」)來表示。因此我們產生四種不同的情緒，其表示如表三：

能量 Energy	壓力 Stress	情緒
快樂 Happy	有活力 Energetic	生氣勃勃的 Exuberance
焦慮 Anxious	有活力 Energetic	焦慮的 Anxious
快樂 Happy	平靜 Calm	令人滿足的 Contentment
焦慮 Anxious	平靜 Calm	沮喪的 Depression

表格 3：情緒模型

2.3 特徵選取

不同時代的作曲者，會因為所處時代不同而有不同的作曲風格，也會受到環境或是文化等因素的影響，因此我們將針對不同時代的作品找出其情緒特徵，並透過特徵選取 (Feature Selection) 選出最佳特徵。特徵選取的目標是要從原有的特徵中挑選出最佳的部分特徵，使其辨識率能夠達到最高值。這些鑑別能力較好的特徵，不但能夠簡化分類器的計算，而且也可以幫助我們瞭解此分類問題的因果關係。其分述如下：

2.3.1 古典音樂特徵

在古典音樂中由於時代以及當時環境的影響，作曲方式較嚴謹，很多作曲家都會承襲特定的風格或是曲式，我們要從古典音樂中找出情緒的特徵，會透過分析旋律所構成的特徵，例如:速度，調性等。而有些特徵(例如:音高、音長)可以利用統計的方式從 MIDI 檔案中取得，有些特徵(例如：調性)則需要透過演算法來計算。我們採用表 4 所列做為古典音樂的情緒特徵

古典音樂特徵	
速度	調(24 大小調)
音高平均	調性(大調小調)
力度平均	拍號

表格 4：古典音樂特徵

2.3.1.1 調性

一首古典音樂作品的調性對於情緒有很大的影響力，例如大調大多有活潑快樂的感覺；小調多有平靜哀傷的感覺，因此調性成為辨識古典音樂作品情緒一個很重要的特徵。本篇論文我們先採用 Krumhansl 等[12]在 1982 年所提出計算調的演算法，其對 24 個調的權重定義如表 5，再由所計算出的調來定義其屬於大調或是小調。其計算方法如下:

$$T = \arg\max \sum_{i=1}^{N} \frac{S^t{}_i}{N} \tag{9}$$

$s^t{}_i$ 表音符對應表 5 所得到的權重

N 表輸入計算音樂片段的音符數目

調	I	II	III	IV	V	VI
大調	6.35	2.23	3.48	2.33	4.38	4.09
小調	6.33	2.68	3.52	5.38	2.6	3.53
調	VII	VIII	IX	X	XI	XII
大調	2.52	5.19	2.39	3.66	2.29	2.88
小調	2.54	4.75	3.98	2.69	3.34	3.17

表格 5：Krumhansl 24 個調對應權重表

計算時需要針對不同的調做移調[5]的步驟，以下面例子說明。

假設今天輸入的音樂片段為 note(60,69)，對應到的音高為 C(Do)和 A(La)，我們必須計算 12 個大小調，如表格 5 所示，以 C、D 大調為例，其運算過程如下：

[5]把音階模式轉入不同的音高

- C 大調：對應到表 6，可找到該音高所對應的值。

Tonality(60,69)->(6.35+3.66)/2

調	C	C#	D	D#	E	F
大調	6.35	2.23	3.48	2.33	4.38	4.09
調	F#	G	G#	A	A#	B
大調	2.52	5.19	2.39	3.66	2.29	2.88

表格 6：C 大調對應權重表

- D 大調：對應到表 7，可找到該音高所對應的值。

Tonality(60,69)->(2.29+5.19)/2

調	D	D#	E	F	F#	G
大調	6.35	2.23	3.48	2.33	4.38	4.09
調	G#	A	A#	B	C	C#
大調	2.52	5.19	2.39	3.66	2.29	2.88

表格 7：C 大調對應權重表

其他的調依此類推，計算完 24 個調之後，分數最高者，則為該調。

2.3.2 流行音樂特徵

流行歌曲也受到時代以及文化的影響，現今的作曲方式比較不會遵守過去嚴謹的作曲形式，作曲的人有些甚至沒有受過正統的音樂訓練，因此我們較難從旋律所構成的特徵來做為辨識情緒的依據。但在編曲形式方面，卻可以從目前台灣編曲的風格找出特定的規則，因此在流行音樂的部份，除了和古典音樂相同的特徵（速度、調性）之外，我們也加入了可以從編曲形式中統計出來的特徵（鼓組出現的時間）。我們採用表 8 所列出的特性，做為流行音樂的情緒特徵。

流行音樂特徵	
速度	鼓組出現的時間
調性	鼓組密度

表格 8：流行歌曲特徵

2.3.2.1 鼓組編曲形式

現今流行音樂可以分為前奏(Intro)、主歌(Verse)、副歌(Chorus)、音樂過門及結尾(Instrumental and Ending)四個部份。一般的歌曲 大多作 AA'BA'的曲式，如圖 5 所示。A 代表主歌，而 B 段是副歌。也就是一首歌的構造是由前奏，兩段主歌 AA' ，一段副歌 B，過門音樂，再重覆到主歌 A' 副歌 B，以及結尾音樂順序地連接而成的。在編曲方面，通常較活潑輕快的歌曲會在 A 就編制鼓組且鼓點較密集；而較平靜哀傷的歌曲會在 B 或是第二次的 A' 才編制鼓組且鼓點較鬆散。因此我們計算鼓組出現第一拍的時間做為特徵，在此我們不考慮歌曲前奏部分，均從 A 段開始計算。

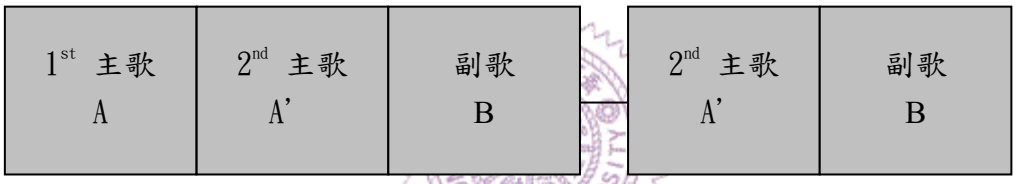

圖 5：流行歌曲曲式

2.3.2.2 歌詞特徵

流行音樂中歌詞可以代表整首歌要給聆聽者的感覺，因此從歌詞中可以找到代表歌曲情緒的特徵。我們先從已分類好的訓練資料中透過斷詞，統計出每個具有情緒的字彙出現的頻率，並透過人工建立同義詞詞表，最後完成情緒詞彙資料庫。

我們所建立的情緒資料庫包含快樂 96 個情緒詞彙(例如:感動、戀愛、快樂等)以及焦慮 223 個情緒詞彙(例如：眼淚、傷心、分離等)。上述情緒詞彙及詞頻對應表見附錄。

2.4 情緒辨識

2.4.1 階層式情緒分類

由於我們採用 Thayer[13]所提出的情緒模型，其使用二維的情緒詞彙組合(「有活力(Energetic)」、「平靜(Calm)」);(「快樂(Happy)」、「焦慮(Anxious)」)，分出四類的情緒，因此在本篇論文中，我們使用階層式的分類方法來進行分類。其表示如圖 6：

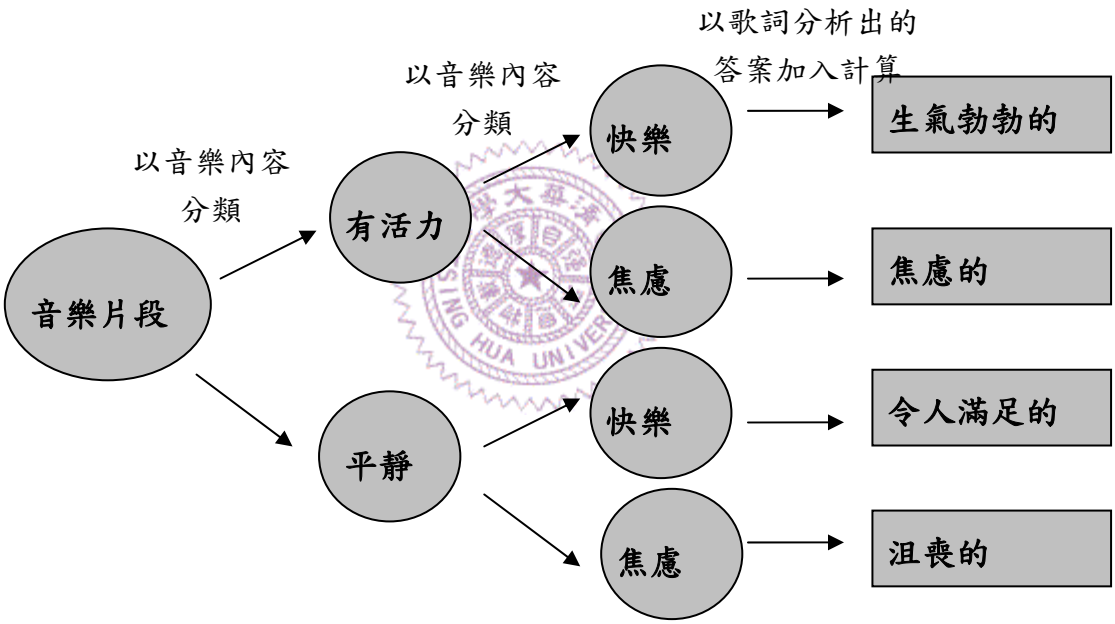

圖 6：階層式情緒分類

2.4.2 歌詞輔助情緒分類

由於歌詞可以表達作詞者想要闡述的意境，因此一首流行歌曲的曲和詞可謂密不可分。我們希望可以結合分析歌詞和分析歌曲所得到的特徵，但因為歌詞部分在「能量」類別區分不明顯，在此我們只將歌詞以「壓力」類別區分，也就是用{‘快樂’}以及{‘焦慮’}兩個詞彙來描述。當有某一類別的權重比例較高時

，代表其重要性較高，而實驗也證明當某一類別有較高的權重比例時，其正確性較高，因此我們設計了下列的計算流程:

1. 計算由歌詞內容所分析出的兩類權重比例相減之值，得到一門檻值。

2. If(測試資料以歌詞內容計算出的分數>=門檻值)

   Answer=歌詞分析計算出的答案

3. else

   Answer=音樂內容計算出的答案

4. 在此門檻值我們透過實驗，使用暴力法找出最佳值。

圖 7 以「下一個永遠」為例，加以說明我們的計算流程。

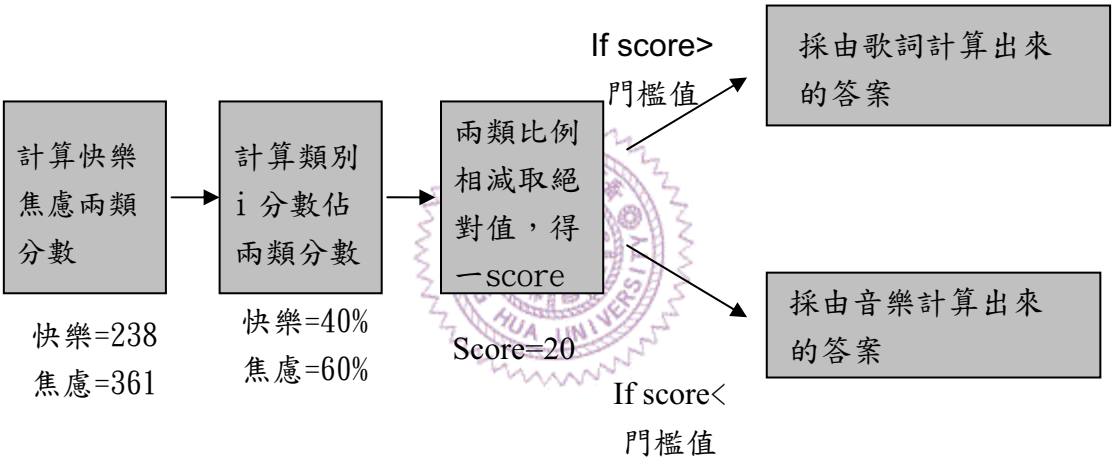

圖 7：歌詞運算流程範例