

Music Localisation: Active Music Content for Web Pages

Dr Ian R O'Keeffe
Centre for Next Generation Localisation,
Localisation Research Centre,
University of Limerick, Ireland

Correspondence: Dr Ian O'Keeffe, Localisation Research Centre, CS1035, CSIS Building, University of Limerick, Co. Limerick, Ireland. E-mail: ian.okeeffe@ul.ie. Phone: 353 61 202781. Fax: 353 61 202734

Abstract

Localisation is a far-reaching discipline, covering much more than just translation. Many other aspects of web site design also require attention, one example being cultural modification. Music content is often included in web sites, but how much thought goes into the cultural suitability of such music? Is the style, the genre, of the music in keeping with the locale of the user? Different cultures can also derive differing meanings and understanding from music due to their cultural conditioning, and an acceptance that the structures of Western Art Music can be viewed as universal is a dangerous assumption to make. For example, some music deemed happy in some cultures can be perceived as melancholy in others. This paper presents a novel approach for possibly 'localising' a piece of music - via a system the author originally created for the purposes of capturing and recreating emotive content in music - thus permitting the cultural modification of musical digital content held in the MIDI (Musical Instrument Digital Interface) file format. A graphical user interface allows users to alter music until they are happy with the resulting cultural or emotive content, and these presets can then be saved and re-applied to any other musical content. This approach has the enormous benefit of allowing all listeners to participate, not just skilled musicians.

Keywords: Localisation; Internationalisation; Emotive Musicology;

1. Introduction

1.1 Research Area

Localisation affects many aspects of online digital content. From a purely translational standpoint, it deals with the translation of the language itself, and the alteration of the text with regard to its font, its directionality, and its layout. But there is more to localisation than just translation, as different cultures react differently to online content, thanks to the cultural conditioning that occurs with every member of the world population as they grow up, and experience life around them. Their expectations, their reactions, and their understanding are shaped by their culture, and what language they speak, how they interact with their peers, etc. The localisation industry attempts to cater for all of these considerations by addressing many aspects of online content that go beyond the language itself and how it is presented.

For instance, colours are an important consideration, as they can have unexpected cultural implications. Take, for example, financial trading, where the colour a currency is presented in can have implications for whether it is in

credit or in debit [Multilingual 2001]; also, white is commonly associated with mourning in Japan [Fact Monster 2004].

Images also need to be handled carefully, as it is dangerous to assume that a symbol on a button, for example, has a clear meaning for all users. Take, for example, the icons for creating numbered or bulleted lists in MS Word. They are designed to appear correct for a left-to-right environment, such as English, and other European languages:

1--- or o---
2--- o---
3--- o---

Most Middle-Eastern languages read right-to-left, and if the icon is to correctly reflect the functionality of its use it should therefore look more like:

---1 or ---o
---2 ---o
---3 ---o

Images can also cause offence in some cultures if used inappropriately, such as the use of images containing hand gestures that may seem innocuous in other cultures [Jacko 2009].

But what about the localisation of digital music content? Firstly, it is necessary to evaluate the existing research in musical cognition, particularly any research that focuses on cross-cultural differences in comprehension and understanding, because if there is no evidence of such cultural diversity then the concept of music localisation becomes superfluous.

The majority of the work in this area tends to focus on the psychological aspects of musical understanding, and the physiological results, rather than on how the music itself may be adapted to suit differing cultures or locales. The research also generally takes the form of passive studies of test subjects and their reactions to pre-prepared musical data, rather than an active approach where the test subjects are able to change the music themselves. Looking firstly at psychological approaches, Gregory and Varney conducted a study looked at the affective response of subjects to music from different cultures [Gregory, Varney 1996], and found that listeners "brought up in the Indian cultural tradition have difficulty in appreciating the emotional connotations of western music". Walker [Walker 1996] states that "understanding the music of another culture requires assimilation of the influences affecting musical

behaviour as much as of the resultant musical products”, suggesting that cultural conditioning plays a part in how a listener understands music. Different cultures and ethnic backgrounds also show preferences for different types of music for musical therapy, as demonstrated in a study of the music selected by medical patients to help with post-operative pain relief [Good et al 2000]. Moving on to neuro-science, a study [Morrison et al 2003] of human brain activity captured using functional magnetic resonance imaging (fMRI) showed that there were activation differences between Western (familiar to test group) and Chinese (unfamiliar) music based on training. Trained listeners showed extra activation “in the right and left midfrontal regions for Western music and Chinese music, respectively”. It would therefore seem that we react differently to unfamiliar musical styles or traditions whether we want to or not!

What is reassuring in the research discussed here is that there does seem to be some evidence that music is not the universal language that it is often thought to be, and that there is a place for musical localisation within the general localisation workflow. Music is fairly pervasive in terms of online web content, in both symbolic (MIDI) and waveform (.wav, .mp3 and so on) file formats, but despite this there seems to be little research in the area of music modification for cultural reasons; this helps in forming an opportunity for proposing a method or system for capturing cultural music templates, and then using these templates to localise musical data. This would then open the possibility of allowing music to adapt automatically to the locale, culture, or even the IP address, of the user.

The localisation process would have to consider two possible paths for musical translation, however, one differing widely from the other in terms of approach.

Firstly, we have the replacement approach, where the process would feature the complete removal of musical content, so it could be replaced with something more culturally suitable. Examples could include attempting to match one form of folk music with something of a similar cultural positioning in another country. This approach would require a huge database of suitable music clips, and a complex management strategy in terms of genre categorization, and cultural suitability, and would run into massive issues of storage, musical rights for performance, and data maintenance. For these reasons, this approach will not be pursued in this paper.

Secondly, there is the adaptation approach, where the existing musical content would be shaped to better match the target culture. This could be viewed as an actual translation of the digital musical content, to make it more culturally-suitable, either emotionally or categorically. This is the approach that this paper will explore here.

Of course, facilitating this kind of flexibility would require the handling of a number of considerations:

- Online musical content would need to be tagged, so that the required response from the user was established in advance. The original author of a webpage, for example, may have a particular mood in mind when selecting a piece of music, but if this is not stored somewhere as metadata then the localisation process becomes one of interpretation, which is always open to error. This tag could represent the emotion of the music required (happy or sad?), or perhaps its category or mood (upbeat, youth culture, news, corporate and so on). This tagging would be similar to the tagging of images with their literal meaning to aid in their localisation, as in

this HTML example, where a *.gif image is tagged as an “Angry face”:

```

```

- Once the required emotion or category was established, then the music would have to be culturally adapted to correlate with the locale or culture of the user, while also maintaining the requirements of the tag. To explain the rationale behind this requirement, consider how cultural conditioning could induce different emotional responses in listeners to the same piece of music dependant on their locale. As an example, consider a news website that tagged a ‘sad’ news story with a piece of music in a minor key. This piece would be accepted as sounding sad to most western listeners, but a very similar scale structure could be regarded as representing joy in some areas of the Middle East, thus causing confusion, and possible anger, if the soundtrack was viewed as a happy or frivolous counterpoint to a tragic news event.
- The next consideration, and it is significant one, is how to alter musical content in the manner we require. The music would need to be stored in some form of notation that would allow it to be edited and parsed, and this notation would preferably be of a fairly universal nature. Then some method of editing the music automatically would need to be put in place. Finally, alteration guidelines suitable for the target culture or context would need to be applied to the piece to correctly ‘localise’ it for the intended user.
- The final consideration, assuming the existence of the prerequisites mentioned above, would be the creation of the musical localisation guidelines themselves. These could be viewed as a form of musical template for the required mood, emotion or category, and their selection and application would facilitate the localisation of the content they are applied to.

What this paper will present is a possible approach for handling all of these considerations, thus providing the localisation industry with a mechanism for automatically, or by choice, altering digital musical content online to match the locale or cultural requirements of the user. In terms of positioning, this approach is best viewed as being placed at the intersection of localisation, music psychology and music technology.

More specifically, this paper is placed within a sub discipline of music psychology, that of music cognition; this area is concerned with the study of music as information, from the viewpoint of cognitive science. This discipline shares the interdisciplinary nature of other fields such as cognitive linguistics. Music technology is the result of applying computers and other forms of technology to the creation and adaptation of music, and localisation here refers to the alteration of musical content to match a locale, a language, or a culture. It should be noted that the initial focus of the research presented here was on the capture of emotional content in music for one particular culture, listeners to western art music, but the data gathered in this particular application demonstrates a lot of promise for the expansion of this technical approach into other areas of music analysis and modification, and it is these potential areas of research that will be covered here.

Expanding the initial study from a western-centric bias to

cover cultural differences in different locales affects two aspects of the research; on the one hand, amending the system to cope with differing musical cultures, such as dealing with different scale and pitch intervals and harmonies, and on the other, conducting studies with the existing system, but with different cultural groups as participants. Such studies could be used to capture templates for individual cultures, with these templates then being used to localise online digital content as it is encountered by the user.

The system presented here uses the MIDI music file format for storage and manipulation of the musical content. It was selected because of its wide availability, portability, and small storage footprint, and because it can be played by almost anything, ranging from PCs to mobile phones and PDAs. It does have some limitations, and perhaps other notation methods would need to be considered in the future, but MIDI is certainly a good place to start given its wide acceptance worldwide.

1.2 Background

The starting point for the research described here was an evaluation of cognitive musicology. Cognitive musicology is an interdisciplinary field of research that evolved during the 1970s from such diverse sources as cognitive anthropology, artificial intelligence, cognitive psychology, linguistics, musicology, neuroscience, psychoacoustics, speech recognition, and semiotics [Laske 1992]. It initially appeared to check all the right boxes where emotive evaluation of musical content was concerned, but its bias towards a computational approach, with the system attempting to mimic human behaviour, was not what was required for collecting the emotive data required. It did, nonetheless, provide useful insights into the manipulation of musical data for the system that was eventually developed.

What was really needed was human interaction so as to enable the extraction of actual cultural indicators from music, and for this reason the research presented here uses human test subjects controlling low-level musical manipulations, and then storing these presets as a representation of their emotive preferences. This positions the system as a research tool, a technical solution for data gathering, rather than as the primary problem solver, and plays to the strengths of both human intellect and computer technology.

When specific study in the area of meaning and emotion in music is considered, mention must be made of Leonard B. Meyer's "Emotion and Meaning in Music" [Meyer 1956]. This work focuses on the meaning and emotion held within music, or more specifically, within Western music's theoretical tradition, as viewed from a psychological perspective. Meyer starts out by isolating a number of contrasting viewpoints on what constitutes musical meaning and the processes involved in communicating it; firstly citing the dichotomy between absolutists, who "insist that musical meaning lies exclusively within the context of the work itself" [p.1], and referentialists, who contend that "music also communicates meanings which in some way refer to the extramusical world of concepts, actions, emotional states, and character" [p.1]. He then points out that these two meanings are by no means mutually exclusive, as they "can and do coexist in one and the same piece of music, just as they do in a poem or a painting" [p.1], conceding an opportunity for establishing some common ground between the opposing arguments. He also makes the concession that his work is primarily focused on the "closed context of the musical work itself" [p.2], but that this does not rule out the importance or existence of other kinds of musical meaning, such as those introduced through cultural reference or cultural

conditioning.

The aim of the research presented here is to postulate how one could capture this cultural content via a new, interactive approach, and then use this data to alter other pieces of music. It is not a musicological study of existing music, but more an empirical, computational approach that involves the subject directly in the process of creating culturally altered music. In some ways, it is analogous to the localisation process of highlighting imagery inappropriate to a certain culture, and altering it to make it more acceptable. When you consider how music is able to affect us physically, bringing us to tears, making us feel overjoyed, or energised, it definitely presents itself as worthy of careful consideration by the designers of online digital content.

The approach presented in this paper strives to give the test subject direct control over the music itself. This approach is realised through the creation and use of an interactive computer system that involves the subject directly in the process of *creating* the required culturally-suitable emotional states or moods through music. This is achieved by crafting a series of low-level musical operations from scratch and creating a system that allows the user to build culturally-specific modification presets with these operations, thus synthesizing the required emotion. This has the enormous benefit of allowing all music listeners to participate, not just skilled musicians, as music affects someone whether they can play an instrument or not.

By contrast, previous approaches to analyzing emotive content in music have involved the test subject as a passive listener, reacting to previously prepared material; or as a skilled musical performer, required to improvise on the spot. The former does not give the test subject any emotional control over the music material, and places the emphasis on description and second-hand reporting, and the latter neglects the emotional input of those not skilled in musical performance. These approaches also do not take cultural influences into consideration, being concerned principally with the search for emotional cues in music generally.

2. The System - Specification Gathering

2.1 Initial phase - prototyping

The initial phase in creating the system was to move through a quick succession of prototypes, each gradually increasing in complexity, to see if the initial skeleton could support the proposed system. Once this was verified, then the system quickly evolved to the stage where it could perform basic musical operations on an inputted music file. This early stage of iterative design proved the plausibility of the technical solution proposed, and allowed the system to enter a more structured phase of development where musical, emotive and technical specifications were gathered and organized in a series of studies.

2.2 Transformations

The first of these was to locate and isolate suitable low-level musical transformations. This focused on the area of musicology, both general and cognitive, as both the general theories of music, and the experiences of others in the sphere of computer modeling, were of relevance. This study resulted in the formulation of a set of musical transformations, grouped in classes such as tempo, pitch and rhythm.

When considering emotional transformation in music, one has to consider the transformations themselves. They must be at the lowest level of musical control, the processes that

alter the musical equivalent of punctuation, spelling and the use of italics in literature, for example. They cannot be of a high enough significance that personal preferences can be supported, or bias imparted. This does not mean that they need be powerless, however. Given a simple musical melody, it is possible to alter it significantly just by changing a few fundamentals; for example, speeding up the tune, changing the instrument playing it, changing the attack of the notes from long and smooth to short and jerky, and so on. By isolating and then grouping all these low level transformations into relevant sets, it should be possible to produce a toolkit, a set of building blocks, for altering the initial piece in any way required. The secret to the significance of the alteration would then be held in the sequence of simple steps, not in the steps themselves, somewhat like the use of a keyboard and a word processor to produce an emotionally-complex literary work despite only being able to add letters and punctuation.

Another important consideration relates to the types of transformations that will be used. The research presented here concerns itself with physically altering the structure of the input melody, of re-composing the piece via these transformations. It is not concerned with performance data, how someone plays a piece, but with the effect of altering a composition to change its emotive content. It does not attempt to map different inter-note intervals, for example, as was done by Bresin and Battel [Bresin & Battel 2000], where five pianists were analysed while playing the first 16 bars of the Andante movement of Mozart's Piano Sonata in G major K545. The research presented here is concerned with alter the piece as it is actually written, not how it is performed.

To initiate the process of compiling a list of transformations it was necessary to break music down to its fundamentals; from this research, a list of possible valid musical transformations was compiled. The final list is shown here:

Tempo, Pitch, Rhythm, Timbre, Harmony, Accompaniment, Dynamics, Drum Rhythm, Attack, Articulation, Scale.

2.3 Emotions

As the target for the system was the creation of emotional change, it was necessary to refer to research relating to human emotions, with a particular focus on what most psychologists refer to as the primary emotions, the basic emotions of which most emotional feeling is made up of. While there were a number of combinations available, the general consensus was towards a fairly well accepted set of emotions, and in the end it simply became a case of selecting the most suitable list, in terms of ease of data comparison, for inclusion in the experiments.

In the main study, it was this list of primary emotions that the users were asked to re-create through use of the system. For localisation purposes, this target list would probably have to be expanded, perhaps involving some aspects of how certain cultures perceive emotion in music, or maybe expanding the categories more into the area of moods; but the main thrust of the conversion would be in making the music, and its implied meaning, fit the locale better. In that respect, primary, or basic, emotions are an eminently suitable starting point given their generally accepted universality. For future work any arbitrary labels may be proposed for template generation, from emotions and moods through to descriptive media categories such as 'news', 'sports', 'up-tempo', 'youthful', and so on.

The list of emotions for the study was compiled after analysing the proposals of several psychologists; Robert Plutchik [Plutchik 1980, 2001], Shand [Shand 1914],

Clynes [Clynes 1980], Izard [Izard 1991], Klaus Scherer [Scherer 1995], and Schopenhauer [Gale 1888]. Of particular interest was Plutchik's classification system, which uses a three-dimensional circumplex model describing the relations among emotion concepts, which are similar to the colours on a colour wheel (Figure 1). The eight sectors are designed to indicate that there are eight primary emotion dimensions defined by the theory arranged as four pairs of opposites.

For this reason, Robert Plutchik's model was selected over those of Shand, Izard and Clynes, as it includes a pairing of opposing emotions, thus allowing further comparisons to be made between emotions as well as analyzing each emotion separately. His contention that more complex emotions are made up of combinations of the more basic ones is also of interest, as perhaps emotive combinations could lead to creating emotive musical content in a similar fashion, such as contempt arising out of a cross between anger and disgust. As a control, an option to model "no emotion" was also added. The list of emotions thus became:

Joy, Sadness, Anger, Fear, Acceptance, Disgust, Surprise, Anticipation and No Emotion.

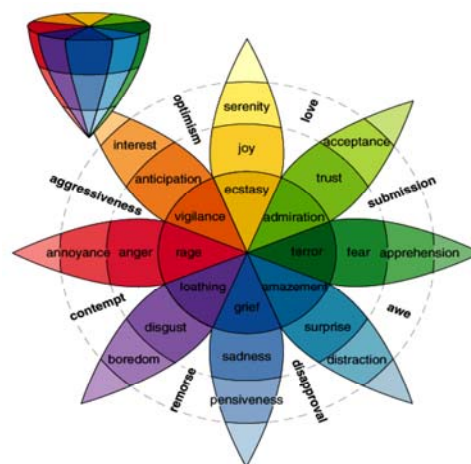


Figure 1 - Plutchik's three-dimensional circumplex model

3. The System - Design

At this stage the system had a requirements list for transformations, and a set of target emotions for users to attain, so work progressed in completing a working first version. Introspective testing, and informal peer reviews, ironed out any major functional issues in the system. A Preliminary Study was conducted, using the primary emotions selected, to help select suitably "unemotional" musical input pieces for use in further testing. This involved composing a list of new, and hence unknown, melodies for review, with those rated lowest in emotional content moving forward for use in later tests. A well known nursery rhyme was also included as a control.

3.1 Development

The development platform was X-Code and Interface Builder on an Apple Macintosh running OSX 10.4.x. Most of the music alteration functionality is written in ANSI C, with the user interface code being handled by Objective C. The system was developed from a small working framework in an iterative manner, using the *spiral design* methodology, with requirements being added as the system grew in size and capability. The advantage of this approach is that it is both incremental and iterative, and allows the development to start out small and benefit from enlightened trial and error throughout the development

process.

The next phase of development was driven by *informal peer reviews* where colleagues tried out the system and highlighted any issues relating to either the user interface layout, or problems with the underlying functionality. From an architecture perspective, the system was now complete.

The final development phase involved users in a *Pilot Study*, which produced results that moved the focus on to improving usability, removing any remaining functional inconsistencies, and fine-tuning the pairing between the physical system and the tasks the participants would be called upon to compete.

3.2 Architecture

The overall system structure consists of three 'layers', named User Interface, Command and Realisation (Figure 3). One layer, the User Interface itself, is developed within Interface Builder, and is solely concerned with tracking, recognising and dealing with on-screen events. These events pass messages to the Command layer, which then fields these messages and decides what action to carry out, and what commands to pass on to the Realisation layer.

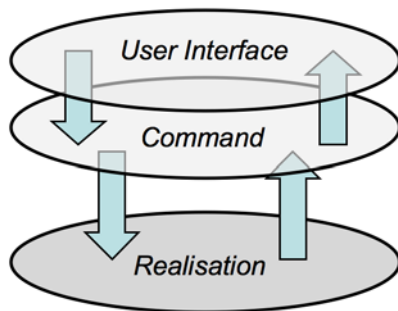


Figure 3 – System Architecture

The Command layer controls the application, listening to the interface for messages that signify button clicks or slider moves, and responds to these events by calling for the relevant action to be carried out. Some button presses require that the lower layer (Realisation) be called; some require management of the appearance of the user interface for controlling the focus of the user, and some deal with playing back sound files, or handling dialogue boxes for opening or saving script files, or inputting MIDI files.

The Realisation layer does all the real work in the system. It controls the input and parsing of the source music file for alteration, storing the musical data into an internal array structure, and then applying whatever transformations the user has requested. It also performs its own housekeeping with respect to sorting, tracking track end points, adding MIDI track headers if absent in the inputted file, and quantisation to ensure all notes are correctly recognised and positioned as they would be in a musical score. Once all the work on the data has been completed it is outputted as a MIDI file for auditioning by the user. A special feature of the Realisation layer is its ability to deal with scripts in a form of batch processing as well as fielding single commands. This allows the user to submit a series of transformations in the form of a 'macro' to any inputted piece, and also means that the developer can create power functions in the Command layer that may use a series of two or more low level functions in combination.

4. The system

4.1 User Guide



Figure 4 – The System

The completed system (Figure 4) enables the user to perform low-level musical manipulations on an imported MIDI (Musical Instrument Digital Interface) file.

The first step involves selecting the MIDI file for transformation, via the "MIDI input from file" button. Once selected, the transformation controls become activated, in vertical columns of similar functions. These columns cover Tempo, Pitch, Rhythm, Timbre, Harmony, Accompaniment, Dynamics, Drum Rhythm, Attack, Articulation, Time Signature, and Scale. The Play and Stop buttons for the original and altered MIDI files remain active throughout the exercise so that the user is able to audition the altered piece after each transformation, and can review the original piece at any time for comparison.

When the user is happy with the emotion he has created for his locale or culture in the resulting piece, he saves the script created for that emotion using "Save Script", and clears the console prior to the next emotion via "Clear Console". Clicking on "Refresh most recent file" clears all transformations from the input file ready for the next pass.

4.2 Scripting

The key to the system's ability to collect emotive data is its scripting functionality. This allows the system to track every decision the user makes in altering the initial melody. However, while simply logging all user actions is a crucial requirement for the research, the ability of the system to re-read the scripts created, and to re-apply them to any new input file, gives it considerably more flexibility. It means that the system can, at a later date, enforce any user's preferences on any inputted melody, or that a generalized script for an emotion, created offline from data gathered using any text editor, can be imported and used separately.

4.3 Evaluation Methods

Analysis of the scripts produced by test subjects was done in an empirical manner using spreadsheets for totaling and averaging the data. The data was compared across emotions, to locate and isolate any noteworthy trends, and also across each participant, to isolate any trends created through participant favouritism rather than emotive effect.

Any scalar controls (tempo, transposition) were tracked in terms of direction as well as magnitude.

The aim of this analysis was to highlight any deviation from what could be regarded as random or normal, and also to compare the divergence between the different emotions, as this would show that the proposed approach had validity for capturing emotive content in music.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Figure 5 - Chi Square Goodness of Fit formula

For the purpose of isolating any significant findings, the Chi Square Goodness of Fit test (Figure 5) was applied to the data, given its nominal/categorical nature. An alpha value of 0.05% was selected for the test for significance, although the sample size may be considered a little on the small side for robust findings at just nine participants. Even so, it gives a good indication of the trends in the data gathered. The two variables that displayed a parametric nature (tempo and pitch change) were analysed to calculate their mean and median values, and their standard deviations.

The collation process created nine tables of data for each participant, one for each emotion tested. These results were then totaled for each emotion, where individual button presses could be selected, or averaged by emotion where integer values could be selected, as for tempo and the pitch change. To isolate personal trends or favouritism by a particular test subject, the data was also totaled for each participant. For example, someone may simply like one option and select it for all emotions, and this data would show this trend and allow it to be accounted for in the overall data analysis.

Profile information was also compiled for each participant, via note taking during the study, informal discussion with the participant during and after the study, and via the questionnaire each participant completed after finishing the study.

Analysis of this profile information produced suggestions for the improvement of the system; and also data that could be cross-referenced with the empirical data, such as the level of musical expertise.

5. Results

5.1 Main Study

The results of the main study show that the idea of collecting emotive data from test subjects using low-level musical transformations definitely has merit, and some interesting trends are apparent even after analyzing the relatively small data sample gathered here.

Tempo is definitely a decisive factor, particularly in Joy, Sadness, Anger and Surprise. What is particularly interesting is that the average tempo for No Emotion is 105bpm, slightly slower than the input file at 120bpm. 120bpm is the default tempo for MIDI; hence its selection for the input file, but the study showed that this was regarded as having "too much" emotion. However, it should be noted that the No Emotion tempo sits quite centrally between the tempi selected for Joy and Sadness, showing its suitability as a median.

Some more examples of emotional opposites in the data, as suggested by Plutchik in his arrangement of the primary emotions as pairs of opposites, can be seen in Pitch between Joy and Sadness, and Anger and Fear; in Rhythm between Joy and Sadness, Acceptance and Disgust, and Surprise and Anticipation; in Dynamics for all emotional pairs; and also in many aspects of Attack Length and Articulation. Scale shows an almost bipolar split between Joy and Sadness, and also shows a lot of variance across the other emotions.

Moving the focus away from each transformation, and onto individual emotional templates, when they are compared in the combined chart (Figure 6), it can be seen that no one emotional template matches another, they are all unique. This demonstrates that each emotion is

definitely mapping to its own set of preferences, and suggesting that there is validity in attempting to extract emotional content from music in this manner. It also demonstrates the strength of an entire set of transformations for an emotion, and hints that an emotional template may well be more than just the sum of its individual transformations.

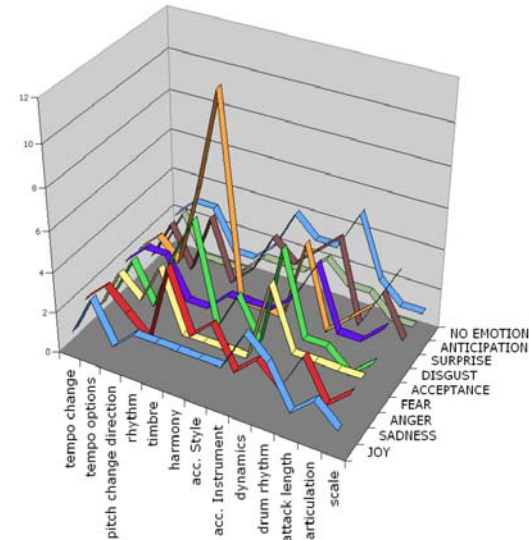


Figure 6 – Emotional Templates Compared

It also tends to imply that a suggested possible result, namely the isolation of a unique combination of low-level musical transformations that could be regarded to represent some form of an emotional template, has been obtained. The close similarity in the results for individual participants for each emotion also lends support to the idea that these templates are demonstrating some consistency, and have some validity in representing their respective emotions. This therefore also implies that using these templates to transform other input pieces into a required emotional state is a viable proposition.

What is also noteworthy is the variance between skilled musicians and those with basic musical abilities in terms of the number of transformations auditioned. The skilled musicians took significantly fewer 'steps' to reach their required goal, but this disparity in the efficiency of the participants was not reflected in the outcome of the study, as it did not affect the close alignment in the data for each emotion. This suggests that the emotions recognised in the music by the participants are not strongly dependant on the level of their musical training, and that the emotional recognition process may be of a much more general cultural nature.

The data therefore suggests that the system has, firstly, been successful in capturing a series of results that differ enough from each other that many of them can be regarded as significant. Secondly, these results are distinct enough that they may be regarded as possible emotional templates. Taking these results, both empirical and observed, into consideration the research can be said to have produced a number of contributions:

- The possibility of extracting empirical data relating to human perception of emotions in music.
- The creation of a system to facilitate this analysis of the human perception of emotion in music, and to also facilitate the alteration of the emotional content of any inputted piece of music.

5.2 Follow-Up Study

The templates produced by the main study presented an excellent opportunity for verifying the data gathered by the system, simply by running the experiment in reverse. To this end, a study was created using a short piece altered emotionally by the system using averaged emotional templates gathered from the main study, and the test subjects were asked to categorise these pieces.

The study involved listening to the 9 pieces - each representing a different emotion - and categorising them by Emotion. No Emotion was included as a control, as before, and users were told that this category should represent an absence of emotion in the piece.

As was the case for the main study, the table compiled by Plutchik defining the eight emotions was supplied to provide the test subject with the same definition of the required target emotions.

The results were split into two groups, those who were already familiar with the system (eight participants), and those who had never seen it (nine participants) and were judging the pieces purely on emotional content (Figures 7 & 8).

The results showed a significant outcome for the data collected for each emotion, although not all emotions have been correctly identified. What is also interesting is how closely the data matches between the two groups, suggesting that there is indeed a recognisable set of emotions in music, whether you are familiar with the system and its processes or just a listener to any emotionally altered piece. The strongest categorisation is for Sadness, with almost all test subjects rating it first. Disgust, Surprise, Joy and No Emotion are very high in certainty, followed by Acceptance and Anger, although some test subjects also linked the Joy piece to Surprise. Fear also produced significant results, although it was sometimes confused with Anticipation. Acceptance was categorized as Joy by the group familiar with the system; in contrast, the independent group classified Acceptance correctly. Anticipation caused widespread confusion, receiving no votes at all from those familiar with the system, and just one vote from the independent test group. Most test subjects categorised it under Joy, with Acceptance and No Emotion being the next most favoured.

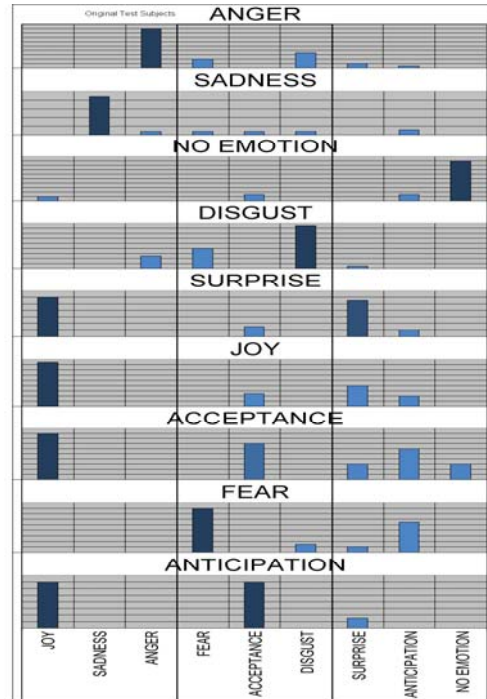


Figure 7 - Data for those familiar with the system

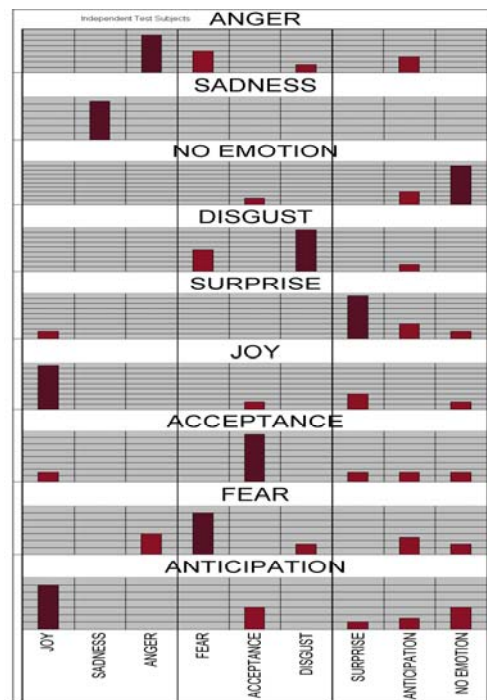


Figure 8 - Data for the independent test subjects

This study essentially closed the loop on the emotional research undertaken, feeding the results back for verification, and demonstrated that pieces generated via the extrapolated emotional templates were, for the most part, correctly identified by a listening test that can be viewed as independent from the Main Study. Some emotions were classified more strongly than others, but where an emotion was correctly identified, statistical calculations suggest that this result should be regarded as significant. The small sample size needs to be taken into account, however, but the early signs are promising for the approach used in this research.

6. Discussion

The results gathered by this study show much promise for an interactive approach to music analysis and alteration. The benefits are observable; clear, empirical data captured from precise on-screen user decisions, avoidance of confusion through bias, introspection, and descriptive issues, ease of usability across a broad musical ability spectrum, and easy capture of emotive templates for re-use. The follow-up study also demonstrated that listeners were able to correctly identify the intended emotion or mood in pieces of music that had been altered by the templates gathered by the system.

To summarise the capabilities the approach provides:

- The ability to capture a user's preferences, in the form of a template, in response to a request to induce an arbitrary mood, emotion or categorization in any piece of music.
- The ability to store this template for re-use on other musical material as required.
- The ability to generalize the data gathered across a number of users for any specific template descriptor; for example, sadness.
- This then implies the ability to segment any generalized findings into any required demographic, such as locale, culture, gender, or whatever.

This, of course, is a general overview of the possible capabilities of the research presented here. What needs to be done next is to consider the possible applications of this functionality.

6.1 Applications for Localisation

Now that the concept of capturing emotive templates for music has been demonstrated, the next phase is to ask how this system could be leveraged in a music localisation environment.

Firstly, from a standards viewpoint, it would be necessary to specify that all symbolic digital musical content for musical localisation is represented in MIDI format, at least until a better format is adopted. Of note are the recent developments in MPEG-4, as it is suggested [Bellini, Nesi & Zoia 2005] that the integration of SMR (symbolic music representation) in the MPEG-4 standard will enable the development of many new applications. Broadening the base of the musical representation would allow more space for handling multimedia opportunities, such as guitar tablature, libretti for operas, and audio-visual components. As of the time of writing, standardization is still in progress. The MPEG AHG (ad hoc group) is pursuing this standardization under the auspices of the MUSICNETWORK, a group funded by the European Commission to help bring music into the interactive multimedia era.

Secondly, a set of categories for the templates would need to be proposed, such as: Basic Emotions (Happy, Sad, Angry, Fearful and so on), or Media Categories (Corporate, News, Sport, Youth Culture, Up-Tempo, and other such terms used within the media industry). The categories and sub-categories would then be used to tag the relevant music content so as to facilitate the localisation process. These categories could always be expanded later as required.

From the perspective of demonstrating the on-demand localisation of existing online music, initial user trials could involve using the existing system to create a number of culturally suitable versions of the musical content in advance, with the website then selecting the correct match

to the user's profile when they connect to the page in question. If a culture was not available, default fallback behaviour could be used, as is currently done for language selection in .NET localisation.

The finished version, however, would be much more sophisticated, having the music transformation routines held on the website servers (possibly in Java for platform independence - as MIDI files, being a symbolic store of performance data, take up very little space and do not place a large load on available processing power), and the specific cultural versions of the required categories and sub-category templates being held on the user's machine as part of their profile in a similar manner to the way in which fonts are currently handled. The web server would then use these user templates to select the transformation functions that matched up to the category tag linked to the musical data, and output a culturally-localised music stream.

Possibly the biggest area of work would relate to the creation of the templates themselves in all the required categories, sub-categories and cultures. There are two main routes to pursue here, either commissioning studies to create musical templates for all required locales or cultures, or by enabling users to create their own templates, with the opportunity to upload these new templates so others can use them also.

The former approach would be similar to commercial localisation practice, would cost money, and would probably lead to a similar divide between commercially viable locales and those that are not regarded as being as lucrative.

The latter could be regarded as being a variant of the crowd-sourcing model, where users would be able to create their own cultural templates for music, possibly with the option of uploading these templates for use by the global community. Then online peer voting could establish the templates viewed as most suitable by the listening community.

The term 'Crowdsourcing' was first coined by Jeff Howe [2006] and later defined as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call." [Howe 2009]. Howe also refers to it as "the future of corporate R&D", citing the example of InnoCentive, the "research world's version of iStockPhoto" [Howe 2006]. What is of more interest here, though, is crowdsourcing motivated simply through personal desire to make a contribution, such as demonstrated by contributors to the translation of Facebook. Their reward is recognition from their peers, and perhaps personal satisfaction. No money changes hands.

What is proposed here is the setting up of a simplified, web-based version of the music modification application that would be open to anyone who wishes to contribute, and create templates for their own particular locale or culture, particularly if they feel the existing templates for emotions, moods or categories do not accurately reflect their beliefs after auditioning music modified by them. In fact, it would be hoped that this form of 'national pride' would provide the motivation for contribution, as it has done for some minority languages on Facebook. Quality enforcement, and the avoidance of online 'vandalism', would be realised by including a peer voting system similar to that implemented by Threadless, the web-based t-shirt company [Brabham 2008]. This would allow visitors to the site to vote on existing templates.

While there would be an initial set of categories and sub-

categories set up (as proposed earlier), it would be wise to allow contributors to also suggest other categories, thus ensuring nothing obvious had been missed. A list of nationalities and languages would also have to be provided, to delineate the templates inputted by culture or locale. Once the initial data was gathered, targeting a much larger user group than in the studies described in this paper, then the true relevance of music localisation could be assessed, as well as the level of demand for such a service.

Further development could see the implementation of these templates in actual websites, and perhaps the facility for the personalisation of the templates on each user's local machine, bringing localisation to possibly its ultimate conclusion, a "locale of one" [Wade 2009], where each user has their own preferences on their own computer for the affective content of all musical content they hear, although the strength of the link to the localisation locale or culture in this instance could be called into question.

In conclusion, this research demonstrates the plausibility of creating a system to capture cultural-specific templates in music by involving the participant directly in the process of creating those templates. While there is still a lot of work to be done from a propagation standpoint, this work presents a new research possibility in the fields of cognitive musicology and localisation - the possibility to perform cultural modification on music for the localisation of online digital musical content - and the system constructed is a strong foundation for the further development of such applications.

Acknowledgements

The author would like to thank his PhD supervisor for her support and advice, and also Enterprise Ireland and the Interaction Design Centre in the University of Limerick for supporting the initial research. This research is also supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cnqli.ie) in the Localisation Research Centre in the University of Limerick.

References

- Bellini, P., P. Nesi and G. Zoia, 2005. Symbolic Music Representation in MPEG. IEEE MultiMedia 2005, Vol. 12, Issue 4: 42 – 49.
- Brabham, D.C., 2008. Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence* 2008; 14; 75
- Bresin, R. and G. U. Battel, 2000. Articulation Strategies in Expressive Piano Performance: Analysis of Legato, Staccato, and Repeated Notes in Performances of the Andante Movement of Mozart's Sonata in G major (K. 545). *Journal of New Music Research*, 29: 211 – 24.
- Clynes, M., 1980. The communication of emotion: theory of sentics. *Theories of Emotion Vol. 1*, R. Plutchik and H. Kellerman (eds), Academic Press, New York, 171 - 216.
- Fact Monster, 2004. What Colors Mean. Pearson Education, Inc. 15 June 2004. <http://www.factmonster.com/ipka/A0769383.html>
- Gale, H., 1888. Schopenhauer's Metaphysics of Music. *New Englander and Yale Review*, Volume 48 Issue CCXVIII, 362-368.
- Good M, B L Picot, S G Salem, C Chin, S F Picot & D Lane, 2000. Cultural Differences in Music Chosen for Pain Relief. *Journal of Holistic Nursing*, Vol. 18, No. 3, September 2000 245-260

Gregory, A. H., N Varney, 1996. Cross-Cultural comparisons in the affective response to music, *Psychology of Music*, 24, 47-52

Howe, J., 2006. 'The Rise of Crowdsourcing', *Wired*, Issue 14.06.

Howe, J., Crowdsourcing: A Definition. Available at: <http://crowdsourcing.typepad.com/> (Accessed November 26, 2009).

Izard, C. E., 1991. *The psychology of emotions*. New York: Plenum.

Jacko, J.A., 2009. *Human-Computer Interaction: Design Issues, Solutions, and Applications*. CRC Press, 2009

Kugel, P., 1992. *Beyond Computational Musicology. Understanding music with AI: Perspectives on music cognition*, MIT Press, 30 – 47.

Laske, O. E., 1992. *Artificial Intelligence and Music: A cornerstone of Cognitive Musicology. Understanding music with AI: Perspectives on music cognition*, MIT Press, 3 – 28.

Meyer, L. B., 1956. *Emotion and Meaning in Music*. Chicago: Chicago University Press.

Morrison, S.J., Demorest, S.M., Aylward, E.H., Cramer, S.C., Maravilla, K.R., 2003. fMRI investigation of cross-cultural music comprehension. *NeuroImage* 20, 378–384.

Nemzow, M., 2001. *Juggling Global Monetary Complexities*. Volume 12 Issue 1 of *MultiLingual Computing & Technology*, Sandpoint, Idaho, USA.

Plutchik, R., 1980. A general psychoevolutionary theory of emotion. *Theories of Emotion Vol. 1*, R. Plutchik and H. Kellerman (eds), Academic Press, New York 171 - 216.

Plutchik, R., 2001. The Nature of Emotions. *American Scientist*, 89: 344 - 350. July/August.

Scherer, K. R., 1995. Expression of Emotion in Voice and Music. *Journal of Voice*, 9(3), 235-248.

Shand, A. F., 1914. *The Foundations Of Character – Being a study of the Emotions and Sentiments*. Macmillan and Co. Limited.

Wade, V., 2009. Supporting a Locale of One. *LRC XIV "Localisation in The Cloud": The 14th Annual Internationalisation and Localisation Conference*, Limerick, Ireland.

Walker, R., 1996. Open peer commentary: Can we understand the music of another culture? *Psychology of Music*, 24, 103-130