TOWARDS USING EXPRESSIVE PERFORMANCE ALGORITHMS FOR TYPIST EMOTION DETECTION

Alexis Kirke
Interdisciplinary Centre for Computer Music Research, University of Plymouth, Plymouth, UK

Matthieu Bonnot
Département Informatique, Polytech'Nantes, Nantes, France

Eduardo R. Miranda
Interdisciplinary Centre for Computer Music Research, University of Plymouth, Plymouth, UK

ABSTRACT

The most common interface for detailed information communication by computer is the QWERTY keyboard. However users have been shown to be notoriously bad at communicating their emotional state when typing emails or SMS messages – hence the reason for so many misunderstood messages, and for netiquette and emoticons. In studies of expressive performance, it has been shown that there are consistencies in the playing styles of pianists and the emotion they are trying to express. There has been initial work on algorithms to emulate such emotional playing styles on computers, and to detect emotion playing styles from MIDI transcripts. These include such elements as tempo and loudness. The emotional playing styles are thought to be related to – amongst other things - the performer's emotional state and the biophysical limitations involved in utilizing a piano keyboard. We propose examining whether such performance emotion detection algorithms can be utilized in examining the emotional state in real-time of a typist using a QWERTY keyboard. Such data would be useful in Human-Computer Interaction studies, and automated emoticon generation systems – amongst others. It also leads naturally to proposals for a protocol for emotional ASCII transmission based on the MIDI protocol.

1. INTRODUCTION

How do humans make their performances sound so different to the so-called "perfect" but robotic performance a computer would give? Two of the most common performance actions found in Western "classical" music are changing the Tempo and the Loudness of the piece as it is played. These should not be confused with the tempo or loudness changes marked in the score, like accelerando or mezzo-forte, but are additional tempo and loudness changes not marked in the score. For example, a common expressive performance strategy is for the performer to slow down as they approach the end of the piece [1]. Another factor influencing expressive performance actions is Performance Context. Performers may wish to express a certain mood or emotion (e.g. sadness, happiness) through a piece of music. Performers have been shown to change the tempo and dynamics of a piece when asked to express an emotion as they play it [2].

Computer Systems for Expressive Music Performance (CSEMPS) are systems designed to add such expression to non-expressive computer performances [3]. One of the most influential CSEMPs is Director Musices. Director Musices (DM) [4] has been an ongoing project since 1982. The output of the equations defines the performance actions. For example the higher the pitch the louder the note is played, or during an upward run of notes, play the piece faster. The results of the equations are added together linearly to get the final performance.

As well as applying humanisation, Director Musices is able to implement emotional expression [4], drawing on work by Gabrielsson & Juslin [5]. Listening experiments were used to define the parameter settings on the DM rules for expressing emotions. Subjects were asked to identify a performance emotion from the list: fear, anger, happiness, sadness, solemnity, tenderness or no-expression. As a result parameters were estimated for each of the 6 rules which mould the emotional expression of a piece. For example for "tenderness": inter-onset interval is lengthened by 30%, sound level reduced by 6dB, and two other rules are used: the Final Ritardando rule (slowing down at the end of a piece) and the Duration Contrast rule (if two adjacent notes have contrasting durations, increase this contrast).

2. ESTIMATING AFFECTIVE STATE

The inverse process of manipulating tempo, loudness and articulation in a performance to generate emotional expression is to estimate the emotional expression within a performance by estimating the values of those parameters. So for example if – relative to what we know is an emotionally "neutral" performance - inter-onset interval lengthens by 30%, and sound level reduces by 6dB, then that performance may be estimated as more Tender (based on the DM experiments above). In fact [6], produced by one of the DM researchers, investigates a fuzzy-rule based system for estimating the emotional content of a performance based on tempo, articulation and loudness. A linear-based model was also utilized in [7] to allow the affective content of a piece of music to be detected based on, amongst other things tempo and articulation.

3. COMPUTER KEYBOARDS AND PIANO KEYBOARDS

Much of the testing of DM work has been on piano sounds. It is interesting to note that the use of a piano keyboard involves (usually) two hands. Furthermore there are many parallels between the structure of music and the structure of text. Music is often broken up into motifs, motifs make up phrases, phrases make up themes and themes make up sections, etc. Text is usually broken up into syllables, which make up words, which make up phrases, then make up sentences, which make up paragraphs that combine into documents. These parallels led us to consider whether the tempo and articulation of a computer typist could provide information about the emotion being expressed (or felt) by the typist, in a similar way to that found in the piano player. There are similarities in the biomechanics of the two processes as well, though a typist usually sits relatively still, and keeps there arms fixed at the wrist; whereas the pianist will move their upper and lower body and arms.

However the similarities suggestive enough to justify investigating the application of methods for estimating emotions in piano performances to estimating emotion expressed by people typing. Such an emotional estimation would have many applications in the area of affective computing [8]. It is well known that text alone is not great at communicating the emotion of the typist – hence the need for emoticons and netiquette [9]. For example when typing email or using a chat room, emotional markers could be generated, or text colour changed to help communicate the emotion behind the typing.

A key point that needs to be clarified is the difference between induced emotion and perceived emotion [10]. For example a listener may enjoy a piece of music like the famous Adagio for Strings by Barber (1938) which most people would describe as a “sad” piece of music. However, because they enjoy it, the induced emotion must be pleasurable (their valence is increased by listening), but the perceived emotion is sadness. Although there are some differences between perceived and felt emotions, perceived and felt emotions are highly correlated [11]. Hence it is not unreasonable to consider that the emotion perceived in a user’s typing may often represent their internal emotional state as well.

Typing speed has been used in EmoteEmail [12] to give extra information to the reader, but interpretations are “more thought having been put into crafting a paragraph” or showing “what parts of the message have been copied and pasted and it might also suggest the attention level of the person writing the message”. Typing speeds have also been used in the past to make facial emotion recognition more accurate [13], however the similarities between typing and piano playing were not observed or utilized. There has also been work in communicating emotion during text chat by having physiological sensors on the typist to detect their emotions based on Galvanic Skin response [14]. Microsoft are investigating the use of pressure sensitive keyboards to adjust font size in chat applications, for emotional expression [15]. Obviously if it possible to detect using solely the keyboard typing features then the need for extra cameras or sensors is removed.

4. INVESTIGATION APPROACH

To investigate the utility of this approach some initial experiments were done. The experimental system is shown in Figure 1.

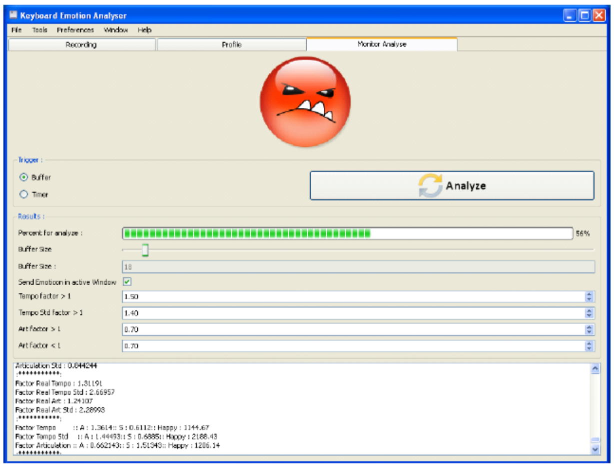

Figure 1. Keyboard Emotion Analyzer

The first observation was that most computer keyboards are not pressure sensitive (despite Microsoft’s recent research) and hence the use of the loudness as part of the emotion detection would require a microphone if the keyboard was not pressure sensitive. Because of this it was decided to focus on timing to simplify matters. On the piano tempo and articulation are calculated as follows. The *onset* X_i is the time at which a piano key is pressed. Assuming monophonic music (as Director Musices’ emotion system assumes) the onsets of a piano key and the next key in seconds are written X_i and X_{i+1}. The *inter-onset interval* (IOI) is defined:

$$IOI_{i+1} = X_{i+1} - X_i$$

Then the tempo of a pair of notes is defined as:

$$Tempo_{i+1} = 1 / IOI_{i+1}$$

The articulation of a note is based on how staccato or legato the note is, and utilizes a similar equation to that used in Director Musices:

$$Articulation_{i+1} = Duration_{i+1} / IOI_{i+1}$$

Duration is how long the piano key is held down for. The lower the articulation, the more staccato a note is. The higher the articulation the more legato it is. From these equations it is possible to find the average values over a number of consecutive piano key presses, and thus calculate the tempo and articulation for groups of notes (e.g. phrases). Translating this to the computer keyboard, X_{i+1} and X_i are the times in seconds of two consecutive key presses. Based on those the inter-onset intervals and articulations can be calculated for keyboard typing as well.

5. INITIAL FEASIBILITY TESTING

A key part of the investigation was to see if estimated music emotion detection parameters from music performances could be used to detect emotion in a typist. To do this a keyboard player was instructed to play three short monophonic tunes using four different emotional expressions: neutral, sad, angry, and happy. For each performance, the tempo and articulation was averaged across the whole performance, and the standard deviation calculated. The emotion clusters Based on these the factors in Table 1 were calculated for the four emotive expression states.

	Tempo Mean	**Tempo Std**	**Articulation Mean**	**Articulation Std**
Normal	1.00	1.00	1.00	1.00
Sad	0.61	0.69	1.06	1.04
Happy	1.72	3.06	0.84	1.75
Angry	2.04	2.02	0.46	1.19

Table 1. Emotion factors

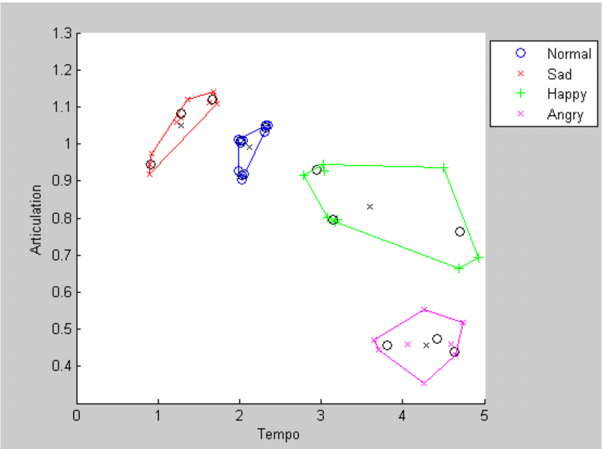

Figure 2. Emotion Clusters

The information contained in the table is interpreted as follow: a sad performance has an average tempo of 61% of normal performances. An angry performance has an average tempo of almost twice as fast as a normal performance. These factors are just initial estimations – as a proper calculation would require a far larger sample. It should also be noted that the use of linear factors is an over simplification, as it can be shown that the factors do not precisely map the normal cluster on to the other clusters. However it was felt that it was sufficient to utilize the linear approach and a single sample to consider the initial feasibility.

Figure 1 shows a system that was developed to implement the factors and emotion detection. It compares the factors in what is being typed live with the emotion factors, and the closest factor is chosen as the detected emotion. This system could work in the background while using such applications as MSN Messenger. If it detected an emotion it would send an emoticon to the active application. Initial tests were run using just tempo. The piano player was asked to type into the computer and comparing their tempo factors to the various tempo factors for different emotions. The typist was asked to type neutrally, angrily, happily and sadly. They would type in messages whose topic was related to the emotion, and try to trigger the emotions internally in themselves, or at least to express them in their typing characteristics. The first thing that was found was that using tempo only meant that happy and angry were hard to distinguish. However happy was distinguishable from sad and neutral.

Another issue that was encountered was the differences in expression between computer keyboards and piano keyboards in terms of articulation. The angry music had levels of staccato which were far sharper than anything found in the piano players typing. This concurs with the authors' intuitive experience – the range of expressive deviations in average computer keyboard use is lower than that found in piano music. To deal with this, a set of expression reduction coefficients were added to compress the size of the expressive factors in table 1. This was found to increase the detection accuracy.

6. NEXT STEPS

Having examined the feasibility of a simple approach, a number of next steps were indicated. These include increasing the sample size to at least 10 musicians. Also the linear factor approach could be extended to a cluster distance measure which takes into account the trained cluster width. Furthermore the initial test only examined tempo mean, the incorporation of articulation mean and tempo and articulation standard deviation could be tried, giving a four dimensional cluster model. Another issue to be addressed is how to improve the testing of the system: is there a way to get typists to type in different emotional ways? Or perhaps a better approach would be to use physiological sensors to detect emotion and look for correlations between cluster distances and emotion.

The approach, if verified, also opens up the possibility of an extended ASCII code which utilizes MIDI in its transmission. Instead of transmitting just ASCII codes for text (as most email and text clients do), a series of ASCII events could be transmitted. So a MIDI file could be transmitted where each note-on event is a key press. This would have the advantage of incorporating the user's typing state into the communication. This could allow emotion detection etc to be done at the other end of the communication.

7. CONCLUSIONS

We have proposed a new approach for detecting the emotion of a typist at a keyboard. Users have been shown to be notoriously bad at communicating their emotional state when typing emails or SMS messages – hence the reason for some many misunderstood messages, netiquette and emoticons. In studies of expressive performance, it has been shown that there are consistencies in the playing styles of pianists and the emotion they are trying to express. There has been initial work on algorithms to emulate such emotional playing styles on computers, and to detect emotion playing styles from MIDI transcripts.

We have proposed examining whether such performance emotion detection algorithms can be utilized in examining the emotional state in real-time of a typist using a QWERTY keyboard. Such data would be useful in Human-Computer Interaction studies, and automated emoticon generation systems – amongst others; it also avoids the need for face detection or physiological sensors in the process used in other studies. Our initial work has looked at utilizing tempo and articulation to indicate emotional state. It was found that articulation needed to be scaled when moving from piano keyboard to computer keyboard; and that some discrimination was possible between three states using tempo: the combined happy or angry state, the sad state and the neutral state. This supports the validity of future work with larger and more controlled samples to investigate and adjust the music algorithms for QWERTY emotion keyboard detection, as well as incorporating staccato and other musical factors such as phrasing.

8. REFERENCES

[1] Friberg, A., Sundberg, J., "Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners", *Journal of Acoustical Society of America* 105, pp1469-1484, 1999.

[2] Gabrielsson, A., Juslin, P., "Emotional expression in music performance: between the performer's intention and the listener's experience", *Psychology of Music* 24, pp68-91, 1996.

[3] Kirke, A., Miranda, E., "A Survey of Computer Systems for Expressive Music Performance", *ACM Computing Surveys* 42(1), 2010.

[4] Friberg, A., Bresin, R., Sundberg, J., "Overview of the KTH rule system for musical performance", *Advances in Cognitive Psychology* 2, pp145-161, 2006.

[5] Gabrielsson, A., Juslin, P., "Emotional expression in music performance: between the performer's intention and the listener's experience", *Psychology of Music* 24, pp68-91, 1996.

[6] Friberg, A., "A fuzzy analyzer of emotional expression in music performance and body motion", *Proceedings of Music and Music Science*, Stockholm, Sweden, 2004.

[7] Kirke, A. Miranda, E. R., "Artificial Social Composition: A Multi-Agent System for Composing Music Performances by Emotional Communication", T. Klouche (Ed.) *Mathematical and Computational Musicology*, Springer, Heidelberg, 2010.

[8] Picard, R., "Affective Computing: Challenges", *International Journal of Human-Computer Studies*, Volume 59(1-2) pp55-64, 2003.

[9] Hanna, B., de Nooy, J., "Negotiating cross-cultural difference in electronic discussion", *Multilingua - Journal of Cross-Cultural and Interlanguage Communication*, 23(3), pp257-281, 2004.

[10] Juslin, P., "From Mimesis to Catharsis: expression, perception and induction of emotion in music", In *Music Communication*, D. Miell, R. Macdonald, D.J. Hargreaves, Eds. Oxford University Press, pp85-116, 2005.

[11] Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., Dacquet, A., "Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts", *Cognition and Emotion* 19, pp1113-1139, 2005.

[12] Angesleva, J., Reynolds, C., O'Modhrain, S., "EmoteMail", *Proceedings of SIGGRAPH*, ACM, Los Angeles, CA, 2004.

[13] Tsihrintzis, G., Virvou, M. Alepis, E., Stathopoulou, I.. "Towards improving visual-facial emotion recognition through use of complementary keyboard-stroke pattern information", *Proceedings of the Fifth International Conference on Information Technology: New Generations*, pp32–37, 2008.

[14] Wang, H., Prendinger, H., Igarashi, T., "Communicating emotions in online chat using physiological sensors and animated text", *Proceedings of CHI - extended abstracts on Human factors in computing systems*, Vienna, Austria, 2004.

[15] Dietz, P., Eidelson, B., Westhues, J., Bathiche, S., "A Practical Pressure Sensitive Computer Keyboard", Report, Microsoft Corporation, 2009.