EXPRESSIVE DIRECTOR: A SYSTEM FOR THE REAL-TIME CONTROL OF MUSIC PERFORMANCE SYNTHESIS

Sergio Canazza, Antonio Rodà, Patrick Zanon

CSC – Center of Computational Sonology
DEI – Dep. of Information Engineering
University of Padua, Italy
ar@csc.unipd.it, {canazza,patrick}@dei.unipd.it
http://www.dei.unipd.it/ricerca/csc/

Anders Friberg

KTH – Royal Institute of Technology
Speech, Music and Hearing
Stockholm, Sweden
andersf@speech.kth.se
http://www.speech.kth.se/music/

ABSTRACT

The Expressive Director is a system allowing real-time control of music performance synthesis, in particular regarding expressive and emotional aspects. It allows a user to interact in real time, for example, changing the emotional intent from happy to sad or from a romantic expressive style to a neutral while it is playing. The Expressive Director was designed in order to merge the expressiveness model developed at CSC and at KTH. The control of the synthesis can be obtained using a two-dimensional space (called “Control Space”) in which the mouse pointer can be moved by the user from an expressive intention to another continuously. Depending on the position, the system applies suitable expressive deviations profiles. The Control Space can be made so as to represent the Valence-Arousal space from music psychology research.

0.1. Definitions

Acoustic parameters: they specify the low level characteristics of a performance; in the case of MIDI performances they are: the onset time O, the inter onset interval IOI, the duration D, the intensity I (expressed in dB), and the note number (which specify the pitch).

Tick: it is a temporal measure, which refers to the nominal duration (specified in the score); equal durations notes in the score have the same duration expressed in ticks.

Metronome (MM): it is a measure of the instantaneous tempo with which a given note is played; it is measured in beats per minute.

Legato degree (leg): it is the ratio between the duration of a given note and the inter onset interval which occurs between its subsequent note: $leg = D/IOI$.

Key Velocity (KV): it is an intensity measure, which is used in the MIDI representation of a musical performance, and it is approximately linearly related to the intensity reproduced by the sound card; its values ranges from 0 to 127.

Sonologic parameters: they specify the mid level characteristics of a performance; in the case of MIDI performances they are: the metronome MM, the legato degree leg, and the Key Velocity KV.

1. INTRODUCTION

The shaping of the musical expression is a natural part of musicians task when performing music. To become a master musician is a time-consuming endeavor to say the least. A great deal of the time is spent mastering the motoric skills in producing the tones. What happens if the note production is the left to the computer but the expressiveness is still controlled by the musician? In order to explore this we developed the Expressive Director, which combines two different approaches for computer modelling of expressiveness in music performance. The starting point was the KTH rule system that contains an extensive set of context-dependent rules going from score to performance processed in non-real-time, and the CSC expressiveness model that transforms a natural performance (rather than the score) and works in real-time. A score, preprocessed in Director Musices, can be played by the Expressive Director with a real-time control of all the rule parameters.

2. GENERAL OVERVIEW OF DIFFERENT REPRESENTATIONS OF THE EXPRESSIVENESS

2.1. The KTH rule system

The KTH rule system is a set of about 30 rules that transform a score to a musical performance [3, 5]. The rule system models different principles used by musicians when performing a given piece. It is intended to model general principles found to be used by many musicians and not restricted to any particular style or instrument, which however, is not always feasible. Another goal has been that a rule should work on any musical context, implying that the context constraints within the rule itself finds the appropriate musical condition where the rule should trigger and filters out the rest. The rules cover such aspects as phrasing, articulation, timing of small groups, intonation, tonal tension, and accents [7].

Each rule has one main parameter k, which is used to control the overall amount of variation. Most rules also have extra parameters to fine adjust the behavior. By combining k values and extra parameters, different performance styles can be obtained. In such a way, it was possible to model variations among musicians playing the same piece [4], or changing the emotional/motional character [1, 8].

The rules have been implemented in a number of different computer programs. The development platform at KTH is the program Director Musices[1] [6] in which most of the rules are implemented.

[1] The software can be downloaded at http://www.speech.kth.se/music/performance

2.2. The CSC expressiveness model

The model is based on the hypothesis, that different expressive intentions can be obtained by suitable modification of a neutral performance. The transformations realized by the model should satisfy some conditions: 1) they have to maintain the relation between structure and expressive patterns found into the neutral performance, 2) they should introduce as few parameters as possible to keep the model simple. In order to represent the main characteristics of the performances, we used only two transformations: shift and range expansion/compression. Different strategies were tested. Good results were obtained [9] by a linear instantaneous mapping that – for each *sonologic parameter* S and for a given expressive intention e – is formally represented by the equation:

$$S_e(n) = k_e \cdot \overline{S}_0 + m_e \cdot \big(S_0(n) - \overline{S}_0\big) \qquad (1)$$

where $S_e(n)$ is the estimated profile of the performance related to expressive intention e for the n-th note, $S_0(n)$ is the value of the S-parameter of the neutral performance, $\overline{S}_0$ is the mean of the profile $S_0(n)$, k_e and m_e are respectively the coefficients of shift and expansion/compression related to expressive intention. We verified that these parameters are very robust in the modification of expressive intentions [10]. Thus, the equation (1) can be generalized to obtain, for each S-parameter, a morphing among different expressive intentions as:

$$S(n) = k(u) \cdot \overline{S}_0 + m(u) \cdot \big(S_0(n) - \overline{S}_0\big) \qquad (2)$$

in which the expressive parameters $k(u)$ and $m(u)$ are not fixed anymore, but depend on the user input and can change from an expressive intention to another continuously. In this way, the problem of the morphing among expressive intention has been translated into the problem of the morphing among the parameters k and m. Section 4.1 will show one of the possible choices (the simplest) in order to solve this task.

2.2.1. The expressive profiles

The arithmetic mean $\overline{S}_0$, used in equations (1) and (2), is calculated over a sliding window whose size can be defined by the user (the context size). It was not calculated over the entire piece, since we found that different phrases requires different strategies for the same expressive rendering. Thus, in the implementation of the expressiveness model we used the following formula:

$$S(n) = k^{(S)}(u) \cdot P^{(S)}_{ave}(n) + m^{(S)}(u) \cdot P^{(S)}_{dev}(n) \qquad (3)$$

where $P^{(S)}_{ave}(n) = \overline{S}_0(w_n)$ is the *expressive profile* of the sonologic parameter average, which is calculated over the window w_n centered around the n-th note; $P^{(S)}_{dev}(n) = S_0(n) - \overline{S}_0(w_n)$ is the expressive profile of the sonologic parameter deviations around the average.

The expressive profiles contain the information related to the neutral performance, and represent all the timing and intensity nuances that the performer used in his interpretation. These nuances are codified into the profiles $P^{(S)}_{ave}$ and $P^{(S)}_{dev}$ which can be used by the model in real time. The sonologic parameter can be $S \in \{MM, leg, KV\}$, thus we have six profiles representing the neutral performance: $P^{(MM)}_{ave}$ $P^{(MM)}_{dev}$ $P^{(leg)}_{ave}$ $P^{(leg)}_{dev}$ $P^{(KV)}_{ave}$ and $P^{(KV)}_{dev}$.

2.2.2. The Expressive Sequencer

The implementation of the CSC's expressiveness model has been made using the EyesWeb[2] platform, a graphical environment for developing multimedia oriented application developed by the Music and Informatics Lab of the University of Genoa [2]. The core of the model has been written into the `ExpressiveSeq`[3] block which is used to synthesize an expressive performance.

The block inputs are: the information of the score, the expressive profiles, and the control parameters (see figure 1). At the first step, the block reads the score and the profiles storing them into memory. Then, if the button Play is pressed, it starts to sequence the MIDI messages.

Four *acoustical parameters* completely specify these MIDI messages: the note number, the onset time O (expressed in ms), the duration D (expressed in ms) and the intensity I which is expressed through the the Key Velocity KV. The note number is provided by the score, while the other parameter values are calculated using the expressive profiles and the control parameters. More precisely, the calculation of the timing parameters can be made with the subsequent formulas:

$$\begin{aligned} O(n+1) &= O(n) + IOI_{[tick]}(n) \cdot C/MM(n) \\ D(n) &= IOI_{[tick]}(n) \cdot C \cdot leg(n)/MM(n) \end{aligned} \qquad (4)$$

where the initial onset time $O(1)$ is fixed to 0 ms, $IOI_{[tick]}$ is the inter onset interval expressed in ticks as stored into the score, and C is a suitable conversion constant; the expressive metronome $MM(n)$, the expressive legato degree $leg(n)$, and the expressive intensity $KV(n)$ are calculated in real time through equation (3) and the values of the six controlling parameters: $k^{(MM)}$ $m^{(MM)}$ $k^{(leg)}$ $m^{(leg)}$ $k^{(KV)}$ and $m^{(KV)}$.

3. INTEGRATION

3.1. From rules to profiles

The integration of the KTH rule model into the CSC's expressiveness system requires that the information regarding the KTH rules is translated into suitable profiles. The KTH rule system adds several deviations to the score. Since the average profiles $P^{(S)}_{ave,N}$ of a nominal performance are mostly constant, then we translated the rules into deviation profiles $P^{(S)}_{dev}$.

Each of the KTH rules can be described in terms of deviations that have to be applied to the score. The deviations are: timing deviations ΔIOI, duration deviations ΔD, and intensity deviations ΔI. The first two quantities are expressed in ms, while the third is expressed in dB. When all the deviations for each rule have been computed, then they are linearly combined, weighted by the respective rule quantities, the so called k_{KTH} parameters.

The timing information can be translated into a profile of metronome deviations. An expression of such a transformation can be derived by considering how the CSC model and the KTH one calculate the IOI, whose expression should produce the same value; for m rules we have:

$$\frac{IOI_{[tick]} \cdot C}{MM_N + \sum_i^m \Delta MM_i} = IOI_N + \sum_i^m \Delta IOI_i \qquad (5)$$

[2] The software can be downloaded at `http://www.eyesweb.org`

[3] The libraries and the patches can be downloaded at `http://www.dei.unipd.it/ricerca/csc/research_groups/mega/mega.html`

where the subscript N indicates values which are related to the the score (nominal values): the symbols $IOI_{[tick]}$, IOI_N, and MM_N refer to the nominal inter onset interval expressed in tick, in ms, and to the nominal metronome respectively. The symbol ΔIOI_i indicates the inter onset interval deviation introduced by the i-th rule, and ΔMM_i is the respective metronome deviation.

In this formula we can immediately see that the property of additivity which holds for the deviations introduced by the KTH rule system, is not directly translated into additivity for the metronome profile. Moreover, the effect of different rules on the sum of metronome deviations cannot be decoupled from each other. However, for small deviations of ΔIOI_i, the equation (5) can be simplified into the following expression for the metronome deviations:

$$\Delta MM_i \simeq -\frac{\Delta IOI_i \cdot MM_N}{IOI_N + \Delta IOI_i} \tag{6}$$

in which the effects of the rules are decoupled, and the additivity property is maintained.

The duration information has to be managed similarly, in order to obtain a profile of legato deviations. More precisely, if we have m rules, the durations computed by the two models are combined in the following expression:

$$\frac{IOI_{[tick]} \cdot C \cdot \left(leg_N + \sum_i^m \Delta leg_i\right)}{MM_N + \sum_i^m \Delta MM_i} = D_N + \sum_i^m \Delta D_i$$

Also in this case, a simplification is needed in order to achieve linearity and decoupling of the effects of different rules: for small values of ΔIOI and ΔD, we can approximate the deviation of legato introduced by the i-th rule with:

$$\Delta leg_i \simeq MM_N \frac{IOI_N \Delta D_i - D_N \Delta IOI_i}{IOI_{[tick],N} \cdot C \cdot (IOI_N + \Delta IOI_i)}. \tag{7}$$

Finally, all the intensities I expressed in dB are translated into a profile of Key Velocities KV. As usual, if we have m rules, then:

$$KV_N + \sum_i^m \Delta KV_i = f\left(I_N + \sum_i^m \Delta I_i\right)$$

where f is a suitable conversion function, which depends on the sound card used. Decoupling and additivity can be achieved by considering that f is approximatively linear, so that:

$$\Delta KV_i \simeq f(\Delta I_i). \tag{8}$$

Thus, for each rule, three profiles are obtained: $P_1 = \Delta MM$, $P_2 = \Delta leg$, and $P_3 = \Delta KV$.

3.2. Relations between parameters

In the KTH rule system, different rules can be weighted by the so called k_{KTH} parameters, allowing them to model performances more closely and to adapt the rules to different situations.

When we converted the KTH rules into profiles, we had to face with the conversion of the controlling parameters too. In fact, we want that the effect of a rule – when its weighting parameter is equal to a certain value – should be the same of the effect of the rule when it is codified into the CSC expressiveness model. To do so, a relation between the controlling parameters has to be found.

This relation can be obtained by considering how generally the controlling parameters acts in the KTH rule system and into the CSC one; to do so we take as references the deviations calculated by the rules and by the formulas (6, 7, 8) when $k_{KTH} = 1$. In the case of the inter onset deviations and the metronome deviations we have:

$$\Delta IOI = k_{KTH} \cdot \Delta IOI\,|_{k_{KTH}=1}$$
$$\Delta MM = k_{MM} \cdot \Delta MM\,|_{k_{KTH}=1}$$

which can be substituted into (6) obtaining:

$$k_{MM} \Delta MM\,|_{k_{KTH}=1} \simeq -\frac{k_{KTH} \Delta IOI\,|_{k_{KTH}=1}\, MM_N}{IOI_N + k_{KTH} \Delta IOI\,|_{k_{KTH}=1}}$$

For small deviations of intr onset intervals, we can neglect their contribution into the denominator, leading to the final formula for k_{MM} (and the other controlling parameters in a similar way):

$$k_{MM} \simeq k_{KTH}, \;\; k_{leg} \simeq k_{KTH}, \;\; k_{KV} \simeq k_{KTH}.$$

Thus, for each rule three profiles are defined which are controlled by three controlling parameters: $k_1 = k_{MM}$, $k_2 = k_{leg}$, and $k_3 = k_{KV}$.

3.3. The Expressive Director

The `ExpressiveSeq` block uses six profiles for the generation of the expressive performances. The number of profiles (and the relative k and m parameters) is strictly what is required by the CSC expressive model, so that no other redundant profile is allowed. More precisely, each sonologic parameter (tempo, legato and intensity) is affected exactly by 2 profiles, the average profile P_{ave} and the deviation profile P_{dev}, see (3). On the other hand each profile P can affect only one sonologic parameter.

When we generalized the model to include the KTH rules, we had to rewrite the block in order to allow more than 6 profiles. This allowed each sonologic parameter to be affected by more than 2 parameters. Thus, the `ExpressiveDirector` was designed to accept more than 6 profiles, each of which can affect any of the 3 sonologic parameters, i.e. each profile P_l can affect more than one sonologic parameter. To do so, the equation (3) has to be rewritten into the subsequent:

$$S = \sum_l^p k_l(u) \cdot \delta_l^{(S)} \cdot P_l$$

where p is the number of profiles, $k_l(u)$ is the controlling parameter for the l-th profile, and the symbol δ_l^S specify if the sonologic parameter S is affected (the value is 1) or not (the value is 0) by the l-th profile. This last information has been included at the beginning of the profiles file.

4. INTERFACES

The control of the Expressive Director is made through the $k_l(u)$ parameters, which depends on the user input. However, the number of parameters is quite large, making difficult handling them. Thus, we developed an interface able to provide a user friendly control of the expressive rendering, allowing the user to change the expressive intention in real time continuously.

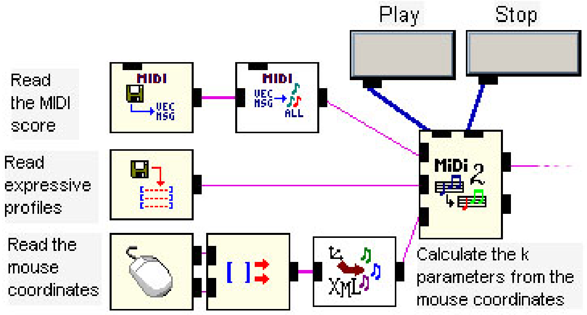

Figure 1: Snapshot of the `ExpressiveDirector` (or `ExpressiveSeq`) at work: the blocks on the left side provide the information needed to synthesize an expressive performance (*score*, *profiles* and k parameters), the central block is the `ExpressiveDirector` (or `ExpressiveSeq`) which control the synthesis, and on the right side, there are the blocks that redirect the output to a MIDI port.

4.1. The control space

The control space controls the expressive content and the interaction between the user and the final expressive performance. In order to realize a morphing among different expressive intentions we developed an abstract control space, called perceptual parametric space (PPS), that is a two-dimensional space derived by multidimensional analysis of perceptual tests on various professionally performed pieces ranging from western classical to popular music [11, 12]. This space reflects how the musical performances might be organized in the listener's mind. It was found that the axes of PPS are correlated to acoustical and musical values perceived by the listeners themselves [13]. We make the hypothesis that a linear relation exists between the PPS axes and each expressive controlling parameter k_l; thus, if x and y are the coordinates of the PPS, then:

$$k_l(x, y) = a_{l,0} + a_{l,1}x + a_{l,2}y.$$

5. CONCLUSIONS

A technical overview of a real time application for expressive rendering of musical performances has been presented, with a brief presentation of the two main contributions that have been merged into the Expressive Director design. Some issues have been examined more deeply, i.e. the translation between the two expressiveness models, both at expressive profiles level, and at the control parameters level.

Some technical refinements have to be carried out:

- implementation into the Expressive Director block of the translation function $KV = f(I)$ in order to be independent from the sound card used and to reduce the approximation of the intensity formulas;
- refinements to the configuration library, in order to manage a large number rules (actually limited at 15).
- assessment of the system has to be made, in order to verify both to which extent the approximated formulas are still valid, and the reliability of the expressive communication channel between the artist using this system and the listeners.

6. ACKNOWLEDGMENTS

This research was supported by the EC project IST-1999-20410 MEGA.

7. REFERENCES

[1] Bresin, R. and Friberg, A. (2000). "Emotional colouring of computer controlled music performance". *Computer Music Journal*, 24(4): 44-62.

[2] Camurri, A., Coletta, P., Peri, M., Ricchetti, M., Ricci, A., Trocca, R., Volpe, G. (2000). "A real-time platform for interactive performance", *Proc. of the ICMC-2000*, Berlin, 374-379.

[3] Sundberg, J., Askenfelt, A. and Frydén, L. (1983). "Musical performance: A synthesis-by-rule approach", *Computer Music Journal*, 7, 37-43.

[4] Friberg, A. (1995). "Matching the rule parameters of Phrase arch to performances of 'Trumerei': A preliminary study", in A. Friberg and J. Sundberg (eds.), *Proceedings of the KTH symposium on Grammars for music performance*, May 27, 1995, pp. 37-44.

[5] Friberg, A. (1995). "A Quantitative Rule System for Musical Expression", *Doctoral dissertation*, Royal Institute of Technology, Sweden.

[6] Friberg, A, Colombo, V, Frydn, L and Sundberg, J (2000). "Generating Musical Performances with Director Musices". *Computer Music Journal*, 24:3, 23-29

[7] Friberg, A. and Battel, G., U. (2002). "Structural Communication". In (R. Parncutt and G. E. McPherson, Eds.) *The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning*. New York: Oxford University Press, 199-218.

[8] Juslin, P. N., Friberg, A., and Bresin, R. (2002). "Toward a computational model of expression in performance: The GERM model". *Musicae Scientiae*, Special issue 2001-2002, 63-122.

[9] Canazza, S., Rodà, A. (1999). "A parametric model of expressiveness in musical performance based on perceptual and acoustic analyses", *Proc. of the ICMC99 Conf.*, November, 1-4.

[10] Canazza, S., De Poli, G., Drioli, C., Rodà, A., Vidolin, A. (2000). "Audio Morphing Different Expressive Intentions for Multimedia Systems", *IEEE Multimedia*, 7(3), 79-84.

[11] Canazza, S., De Poli, G., A., Vidolin, A. (1997). "Perceptual Analysis of the Musical Expressive Intention in a Clarinet Performance". In (M. Leman ed.) *Music, Gestalt, and Computing*, Springer Verlag, 441–450.

[12] Canazza, S., Orio, N., (1999). "The Communication of Emotions in Jazz Music: a Study on Piano and Saxophone Performances", In (Marta Olivetti Belardinelli ed.) *Musical Behaviour and Cognition*, 263–278.

[13] Canazza, S., De Poli, G., A., Vidolin, A. (1996). "Perceptual analysis of the musical expressive intention in a clarinet performance", *IV International Symposium on Systematic and Comparative Musicology*, September, Brugge, 31–37.