Playing with cases: Tempo Transformations of Jazz Performances using Case-Based Reasoning

Ramon Lopez de Mantaras, Maarten Grachten, Josep-Lluis Arcos

IIIA, Artificial Intelligence Research Institute
CSIC, Spanish National Research Council
Campus UAB, 08193 Bellaterra, Spain
mantaras@iiia.csic.es, maarten@iiia.csic.es, arcos@iiia.csic.es

Abstract

The research described here focuses on global tempo transformations of monophonic recordings of saxophone jazz performances. We have investigated the problem of how a performance played at a particular tempo can be automatically rendered at another tempo while preserving its expressivity. That is, listeners should not be able to notice, from the expressivity of a performance, that has been scaled up or down from another tempo. To do so, we have developed a case-based reasoning system called TempoExpress. The results we have obtained have been extensively compared against a standard technique called Uniform Time Stretching (UTS), and we show that our approach is superior to UTS.

Introduction

It has long been established that when humans perform music the result is never a literal rendering of the score. As far as the performed deviations are intentional they are commonly thought of as conveying expressivity. The field of expressive music research comprises a rich and heterogeneous number of studies. Some are aimed at verbalizing knowledge of musical experts on expressive music performance. For example, Friberg et al. have developed Director Musices (DM), a system that allows for automatic rendering of MIDI scores (Friberg et al. 2000). DM uses a set of expressive performance rules that have been formulated with the help of a musical expert using an analysis-by-synthesis approach (Sundberg, Friberg, & Fryden 1991). Widmer (Widmer 2000) has used machine learning techniques like Bayesian classifiers, decision trees, and nearest neighbor methods, to induce expressive performance rules from a large set of classical piano recordings. In another study by Widmer (Widmer 2002), the focus was on discovery of simple and robust performance principles rather than obtaining a model for performance generation. Hazan et al. (Hazan et al. 2006) have proposed an evolutionary generative regression tree model for expressive rendering of melodies. The model is learned by an evolutionary process over a population of candidate models. In the work of Desain and Honing and[1] co-workers, the focus is on the cognitive validation of computational models for music perception and musical expressivity. They have pointed out that expressivity has an intrinsically perceptual aspect, in the sense that one can only talk about expressivity when the performance itself defines the standard (e.g. a rhythm) from which the listener is able to perceive the expressive deviations (Honing 2002). In more recent work, Honing showed that listeners were able to identify the original version from a performance and a uniformly time stretched version of the performance, based on timing aspects of the music (Honing 2006). Timmers et al. have proposed a model for the timing of grace notes that predicts how the duration of certain types of grace notes behaves under tempo change, and how their durations relate to the duration of the surrounding notes (Timmers et al. 2002). A precedent of the use of a case-based reasoning system for generating expressive music performances is the SaxEx system (Arcos, Lopez de Mantaras, & Serra 1998; Lopez de Mantaras & Arcos 2002). The goal of SaxEx is to generate expressive melody performances from an inexpressive performance, allowing user control over the nature of the expressivity, in terms of expressive labels like 'tender', 'aggressive', 'sad', and 'joyful'. Another case-based reasoning system is Kagurame (Suzuki 2003). This system renders expressive performances of MIDI scores, given performance conditions that specify the desired characteristics of the performance. Although the task of Kagurame's system is performance generation, rather than performance transformation (as in the work presented here), it has some sub tasks in common with our approach, such as performance to score matching, segmentation of the score, and melody comparison for retrieval. Recently, Tobudic and Widmer (Tobudic & Widmer 2004) have proposed a case-based approach to expressive phrasing, that predicts local tempo and dynamics and showed it

outperformed a straight-forward k-NN approach. An important issue when performing music is the effect of tempo on expressivity. It has been argued that temporal aspects of performance scale uniformly when tempo changes (Repp 1994). That is, the durations of all performed notes maintain their relative proportions. This hypothesis is called relational invariance (of timing under tempo changes). However, counter-evidence for this hypothesis has been provided (Desain & Honing 1994; Friberg & Sundstr¨om 2002; Timmers et al. 2002), and a recent study shows that listeners are able to determine above chance-level whether audio-recordings of jazz and classical performances are uniformly time stretched or original recordings, based solely on expressive aspects of the performances (Honing 2006). Our approach also experimentally refutes the relational invariance hypothesis by comparing the automatic transformations generated by TempoExpress against uniform time stretching.

TempoExpress

Given a MIDI score of a phrase from a jazz standard, and given a monophonic audio recording of a saxophone performance of that phrase at a particular tempo (the source tempo), and given a number specifying the target tempo, the task of the system is to render the audio recording at the target tempo, adjusting the expressive parameters of the performance to be in accordance with that tempo. TempoExpress solves tempo transformation problems by case-based reasoning. Problem solving in case-based reasoning is achieved by identifying and retrieving a problem (or a set of problems) most similar to the problem that is to be solved from a case base of previously solved problems (also called cases), and adapting the corresponding solution to construct the solution for the current problem. To realize a tempo transformation of an audio recording of an input performance, TempoExpress needs an XML file containing the melodic description of the recorded audio performance, a MIDI file specifying the score, and the target tempo to which the performance should be transformed (the tempo is specified in the number of beats per minute, or BPM). The result of the tempo transformation is an XML file containing the modified melodic description, that is used as the basis for synthesis of the transformed performance. For the audio analysis (that generates the XML file containing the melodic description of the input audio performance) and for the audio synthesis, TempoExpress relies on an external system for melodic content extraction from audio, developed by Gomez et al. (Gomez et al. 2003b). This system performs pitch and onset detection to generate a melodic description of the recorded audio performance, the format of which complies with an extension of the MPEG7 standard for multimedia content description (Gomez et al. 2003a). We apply the edit-distance (Levenshtein 1966) in the retrieval step in order to assess the similarity between the cases in the case base (human performed jazz phrases at different tempos) and the input performance whose tempo has to be transformed. To do so, firstly the cases whose performances are all at tempos very different from the source tempo are filtered out. Secondly, the cases with phrases that are melodically similar to the input performance (according to the edit distance) are retrieved from the case base. The melodic similarity measure we have developed for this is based on abstract representations of the melody (Grachten, Arcos, & Lopez de Mantaras 2005) and has recently won a contest for symbolic melodic similarity computation (MIREX 2005). In the reuse step, a solution is generated based on the retrieved cases. In order to increase the utility of the retrieved material, the retrieved phrases are split into smaller segments using a melodic segmentation algorithm (Temperley 2001). As a result, it is not necessary for the input phrase and the retrieved phrase to match as a whole. Instead, matching segments can be reused from various retrieved phrases. This leads to the generation of partial solutions for the input problem. To obtain the complete solution, we apply constructive adaptation (Plaza & Arcos 2002), a reuse technique that constructs complete solutions by searching the space of partial solutions. The solution of a tempo-transformation consists in a performance annotation. This performance annotation is a sequence of changes that must be applied to the score in order to render the score expressively. The result of applying these transformations is a sequence of performed notes, the output performance, which can be directly translated to a melodic description at the target tempo, suitable to be used as a directive to synthesize audio for the transformed performance. To our knowledge, all the performance rendering systems mentioned in the introduction deal with predicting expressive values like timing and dynamics for the notes in the score. Contrastingly, TempoExpress not only predicts values for timing and dynamics, but also deals with more extensive forms of musical expressivity, such as note insertions, note deletions, consolidations of several notes into a long single note, fragmentations of a single note into several shorter ones, and ornamentations.

Results

In this section we describe results of an extensive comparison of TempoExpress against uniform time stretching (UTS), the standard technique for changing the tempo of audio recordings, in which the temporal aspects (such as note durations and timings) of the recording are scaled by a constant factor proportional to the tempo change. The results of both tempo transformation approaches have been evaluated by comparing them to the performances of a professional musician. More specifically, let MS be a melodic description of a

performance of a given musical phrase by a musician at the source tempo S and let MT be a melodic description of a performance of the same musical phrase at the target tempo T by the same musician. Using TempoExpress (TE) and UTS we derived, from MS, two melodic descriptions, MTE and MUTS, at the target tempo T. Next we evaluated both derived descriptions by computing the distance to the target description MT using a distance measure, that was implemented as an edit-distance that computes the difference between the sequences of notes in melodic descriptions. The parameters in the distance measure were optimized using the results of a web-survey in which human subjects rated the perceived dissimilarity between different performances of the same melodic fragment. In this way, the results of TempoExpress and UTS were compared on 6364 tempo-transformation problems, using 64 different melodic segments from 14 different phrases. The results show an increasing distance to the target performance with increasing tempo change (both for slowing down and for speeding up), for both tempo transformation techniques. This is evidence against the hypothesis of relational invariance mentioned earlier. Secondly, we observed a remarkable effect in the behavior of TempoExpress with respect to UTS, which is that TempoExpress improved the results of tempo transformation specially when slowing performances down. When speeding up, the distance to the target performance stayed around the same level as with UTS. In the case of slowing down, the improvements with respect to UTS were statistically very significant (the p-values of Wilcoxon's signed-rank test are smaller than 0.001 for tempos which are between 20% and 70% slower than the source tempo). The p-values are rather high for tempo change ratios close to 1, meaning that for very small tempo changes, the difference between TempoExpress and UTS is not statistically significant. This is in accordance with the common sense that slight tempo changes do not require many changes. In other words, relational invariance approximately holds when the tempo changes are very small.

Conclusions

In this paper we have summarized our research results on a case-based reasoning approach to global tempo transformations of music performances. We have addressed the problem of how a performance played at a particular tempo can be automatically rendered at another tempo preserving expressivity. We focused our study in the context of standard jazz themes and, specifically on saxophone jazz recordings. Moreover, we have briefly described the results of an extensive experimentation over a case-base of more than six thousand transformation problems. TempoExpress clearly performs better than UTS when the target problem is slower than the source tempo. When the target tempo is higher than the source tempo the improvement is much less significant. Nevertheless, TempoExpress behaves as UTS except in transformations to really fast tempos. However, this result is not surprising because of the lack of example cases with very fast tempos.

References

Arcos, J. L.; Lopez de Mantaras, R.; and Serra, X. 1998. Saxex : a case-based reasoning system for generating expressive musical performances. *Journal of New Music Research* 27(3):194–210.

Desain, P., and Honing, H. 1994. Does expressive timing in music performance scale proportionally with tempo? *Psychological Research* 56:285–292.

Friberg, A., and Sundstr¨om, A. 2002. Swing ratios and ensemble timing in jazz performance: evidence for a common rhythmic pattern. *Music Perception* 19(3):333–349.

Friberg, A.; Colombo, V.; Fryd´en, L.; and Sundberg, J. 2000. Generating musical performances with Director Musices. *Computer Music Journal* 24(1):23–29.

Gabrielsson, A. 1987. Once again: The theme from Mozart's piano sonata in A major (K. 331). A comparison of five performances. In Gabrielsson, A., ed., *Action and perception in rhythm and music*. Stockholm: Royal Swedish Academy of Music. 81–103.

Gabrielsson, A. 1995. Expressive intention and performance. In Steinberg, R., ed., *Music and the Mind Machine*. Berlin: Springer-Verlag. 35–47.

Gomez, E.; Gouyon, F.; Herrera, P.; and Amatriain, X. 2003a. Using and enhancing the current MPEG-7 standard for a music content processing tool. In *Proceedings of Audio Engineering Society*, 114th Convention.

Gomez, E.; Grachten, M.; Amatriain, X.; and Arcos, J. L. 2003b. Melodic characterization of monophonic recordings for expressive tempo transformations. In *Proceedings of Stockholm Music Acoustics Conference* 2003.

Grachten, M.; Arcos, J. L.; and Lopez de Mantaras, R. 2004. TempoExpress, a CBR approach to musical tempo transformations. In *Advances in Case-Based Reasoning. Proceedings of the 7th European Conference*, ECCBR 2004, Lecture Notes in Computer Science. Springer.

Grachten, M.; Arcos, J. L.; and Lopez de Mantaras, R. 2005. Melody retrieval using the Implication/Realization

model. *MIREX* http://www.music-ir.org/evaluation/mirex-results/articles/similarity/grachten.pdf.

Grachten, M.; Arcos, J. L.; and Lopez de Mantaras, R. 2006. A case based approach to expressivity-aware tempo transformation. *Machine Learning Journal* 65(2-3):411-437.

Hazan, A.; Ramirez, R.; Maestre, E.; Perez, A.; and Pertusa, A. 2006. Modelling expressive performance: A regression tree approach based on strongly typed genetic programming. In et al., R., ed., *Proceedings on the 4th European Workshop on Evolutionary Music and Art*, 676–687.

Honing, H. 2002. Structure and interpretation of rhythm and timing. *Tijdschrift voor Muziektheorie* 7(3):227–232.

Honing, H. 2006. Is expressive timing relational invariant under tempo transformation? *Psychology of Music*. (in press).

Juslin, P. 2001. Communicating emotion in music performance: a review and a theoretical framework. In Juslin, P., and Sloboda, J., eds., *Music and emotion: theory and research*. New York: Oxford University Press. 309–337.

Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady* 10:707–710.

Lopez de Mantaras, R., and Arcos, J. L. 2002. AI and music: From composition to expressive performance. *AI Magazine* 23(3):43–58.

Palmer, C. 1996. Anatomy of a performance: Sources of musical expression. *Music Perception* 13(3):433–453.

Plaza, E., and Arcos, J. L. 2002. Constructive adaptation. In Craw, S., and Preece, A., eds., *Advances in Case-Based Reasoning*, number 2416 in Lecture Notes in Artificial Intelligence. Springer-Verlag. 306–320.

Repp, B. H. 1994. Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study. *Psychological Research* 56:285–292.

Sloboda, J. A. 1983. The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology* 35A:377–396.

Sundberg, J.; Friberg, A.; and Fryd´en, L. 1991. Common secrets of musicians and listeners: an analysis-by-synthesis study of musical performance. In Howell, P.; West, R.; and Cross, I., eds., *Representing Musical Structure*, Cognitive Science series. Academic Press Ltd. chapter 5.

Suzuki, T. 2003. The second phase development of case based performance rendering system “Kagurame”. In Working Notes of the IJCAI-03 Rencon Workshop, 23–31. Temperley, D. 2001. *The Cognition of Basic Musical Structures*. Cambridge, Mass.: MIT Press.

Timmers, R.and Ashley, R.; Desain, P.; Honing, H.; and Windsor, L. 2002. Timing of ornaments in the theme of beethoven’s paisiello variations: Empirical data and a model. *Music Perception* 20(1):3–33.

Tobudic, A., and Widmer, G. 2004. Case-based relational learning of expressive phrasing in classical music. In *Proceedings of the 7th European Conference on Case-based Reasoning* (ECCBR’04).

Widmer, G. 2000. Large-scale induction of expressive performance rules: First quantitative results. In *Proceedings of the International Computer Music Conference* (ICMC2000). San Francisco, CA: International Computer Music Association.

Widmer, G. 2002. Machine discoveries: A few simple, robust local expression principles. *Journal of New Music Research* 31(1):37–50.