



**HAL**  
open science

# ESP: Roadmaps as Constructed Interpretations and Guides to Expressive Performance

Elaine Chew, Liu Jie, Alexandre R J Francois

► **To cite this version:**

Elaine Chew, Liu Jie, Alexandre R J Francois. ESP: Roadmaps as Constructed Interpretations and Guides to Expressive Performance. 1st ACM workshop on Audio and music computing multimedia (AMCMM '06), Oct 2006, Santa Barbara, California, United States. 10.1145/1178723.1178744 . hal-03854704

**HAL Id: hal-03854704**

**<https://hal.science/hal-03854704>**

Submitted on 16 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ESP: Roadmaps as Constructed Interpretations and Guides to Expressive Performance

Elaine Chew and Jie Liu

Epstein Department of Industrial and Systems Eng  
Integrated Media Systems Center  
University of Southern California Viterbi Sch of Eng  
Los Angeles, California 90089-0193, USA

{echew, jliu}@usc.edu

Alexandre R. J. François

Computer Science Department  
Viterbi School of Engineering  
University of Southern California  
Los Angeles, California 90089-0273, USA

afrancoi@usc.edu

## ABSTRACT

The Expression Synthesis Project (ESP) literalizes the driving metaphor to allow users to create expressive performances through a driving interface. In musical performance, controlled variations in tempo, dynamics, and articulation, lead to musical gestures that clarify structural ambiguities and create musical affect. Guided by the road in ESP, the user changes the tempo by depressing the gas or brake pedals; dynamics are coupled with acceleration, and articulation is controlled independently. Straight road segments suggest opportunities to accelerate, while varying degrees of road bends prompt slowdowns of different levels. The road serves as a representation of, and guide to, a particular interpretation of a piece. This paper focuses on the concepts underlying the design of the road in ESP, providing case studies on mapping diverse interpretations to roadmaps. We describe the design of two roadmaps each for Brahms' *Hungarian Dances Nos. 2 and 3*. We show that the road has a distinct influence on the way a driver traverses a path, presenting empirical results of driving/performance outcomes using the roads designed, and analyzing the tempo time series. General trends of slowdowns near curvature extremities and accelerations are observed. Statistics show that a particular roadmap has the effect of constraining the performances to a certain style, with specifically placed gestures, while allowing the driver latitude in expressive control. Performances from the same paths are shown to be more similar to each other than those from different paths, having average correlation coefficients of 0.39 vs. 0.16 for Dance No.2, and 0.61 vs. 0.38 for Dance No.3. Overall, the interpretation maps resulted in expressively pleasing performances made available for further examination.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AMCMM'06, October 27, 2006, Santa Barbara, California, U.S.A.

Copyright 2006 ACM 1-59593-500-2/06/0010...\$5.00.

## Categories and Subject Descriptors

H.5.5 [Information Interfaces and Presentation]: Sound and Music Computing – *methodologies and techniques*; H.5.2 [Information Interfaces and Presentation]: User Interfaces – *input devices and strategies*.

## General Terms

Design, Experimentation, Human Factors, Performance.

**Keywords:** Expression synthesis, driving interface, expressive performance, musical interpretation, motion metaphor.

## 1. INTRODUCTION

This paper focuses on musical interpretation and the encoding of an interpretation in the mapping of a road within the Expression Synthesis Project's (ESP's) driving interface for generating expressive performances. The ESP driving interface is shown in Figure 1.

Musical interpretation has been the subject of much research (see [7]). When a composer creates and documents a composition for subsequent re-creation, the written score is an approximation of the idea of the piece. The approximation is incomplete, leaving

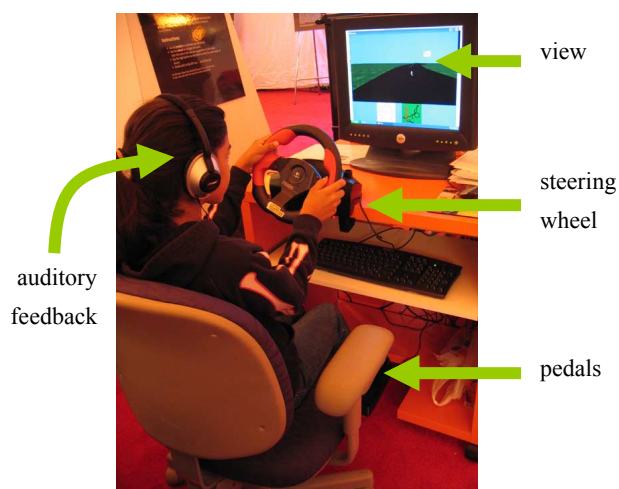


Figure 1. The ESP driving interface for creating expressive performances

ample room for interpretation when a performer realizes the score in a performance. Through the shaping of tempo, dynamics, and articulation in the performance of a musical piece, an interpretation serves to convey musical meaning by clarifying structural ambiguities, communicating expressive gestures, and evoking emotional effect (see p.118-121 in [11]). An interpretation is the result of analysis, as well as decisions on gesture and affect. A well-crafted interpretation can focus attention, create interest, and clarify important structural patterns in the music. It is important to point out that there is not just one solution to an interpretation of a score. There exist multiple interpretations of viable and effective grouping boundaries, and patterns of tension, relaxation, and motion, in the reading or hearing of any composition.

ESP [2, 4], takes the metaphor of driving for expressive performance literally to create a driving interface for the control of expressive parameters such as tempo, loudness, and articulation. Anecdotal evidence suggests that performing a piece of music also shares many similarities with the act of driving a car. Motion in an automobile possesses the added feature of machine actuation of acceleration and deceleration. The advantage of a literal reading of the metaphor of driving for expressive performance is that in both driving and expressive control, smooth tempo (speed) control is a hallmark of an expert operation. The ESP interface allows novices to create expressive performances without having to master an instrument, it allows performers to experiment with interpretations prior to mastering the notes in a piece, and it provides a means for composers who may not be expert performers on numerous instruments, to record their preferred interpretation of their compositions for these instruments.

The focus of this paper is the mapping of musical interpretations systematically to the road in the virtual environment in ESP, so as to guide the user toward particular expressive choices. Todd [13] has shown that crescendo (increasing loudness) and decrescendo (decreasing loudness) shapes indicate structural groups, and the level of embedding determines the dynamic range; in addition, dynamics are coupled to the tempi. Therefore, one can infer from these observations (and experience) that tempo modulation is an important tool in clarifying structural ambiguities. In ESP, we exploit patterns of driving behavior (such as slowing down at a bend to maintain control and speeding up when the road is straight) to guide the user towards particular expressive gestures that can help resolve certain structural ambiguities.

In ESP, the user manipulates tempo and dynamics parameters by depressing the gas or brake pedals. While the function of the pedals in the ESP interface is clear, the wheel and road play an equally important, although less immediately apparent, role in expression synthesis. By mapping an interpretation, a particular sequence of motion patterns and grouping boundaries, to a road that suggests certain driving patterns, we use the road to influence the performance of a piece of music. The user's actions are guided by the road in the virtual environment. Straight segments suggest opportunities to accelerate, while varying degrees of road bends prompt different levels of slowdowns. It is the steering wheel and the road that allows the user to interact with the road designed so as to create a pleasing performance. It is important to note that the road is simply a suggestion of an interpretation. Nothing stands in the way of the user who wishes to accelerate

into, and speed through, a curve, nor one who decides to slow down during a straight section. As in live expressive performance on a traditional instrument, the choice is entirely up to the user/performer. In the case of the ESP interface, the road not only serves to guide the expression, the resulting map represents a particular interpretation of the piece.

This paper introduces the concepts underlying the design of the road, and case studies on mapping interpretations to roadmaps and vice versa. The purpose of this paper is to make clear the connections between the road and a particular interpretation, and to show that such a strategy indeed guides the user towards particular expressive interpretations over others. The remainder of the paper is organized as follows: in Section 2, we demonstrate the mapping of two interpretations each to Brahms' *Hungarian Dances Nos. 2 and 3*. We show the roadmaps constructed from each of these four interpretations of the two pieces. In Section 3, we show and analyze empirical data collected from the "performances" created on these roads in ESP. Graphs of tempo/acceleration over musical time (fractions of a beat) are compared to the corresponding curvature graphs and analyzed for similarity across performances on different roadmaps. We present conclusions and discussions in Section 4.

## 1.1 About ESP

In this section, we quickly outline the capabilities and mappings of the ESP interface. The details of the system design and mapping strategies can be found in [2], [5] and [9]. In ESP, the pedals control the speed of the car. Like most driving interfaces, depressing the right pedal causes the car to speed up. Depressing the left pedal activates the brakes, and causes the car to slow down. Tempo control on a prescribed road is a focus of this paper. The system design and the software architecture style for enabling real-time interaction is described in [5]. The mapping strategies to ensure tempo smoothness are described in detail in [9].

The road, the other central topic of this paper, represents the music, and one particular interpretation of its grouping structure. Prior research has shown that tempo slowdowns tend to occur at group boundaries, for example, at the edge of phrases. These interpretative cues are embedded in the twists and turns in the road. Bends in the road suggest places to slow down, and straight sections encourage speedups. By careful placement of these bends and straight sections, we design of the road in such a way as to guide the user towards a particular, hopefully desirable, interpretation.

The car can be thought of as a pointer to the music file. Fast travel on the road results in a fast tempo replay, and vice versa for slower speeds. Extra buttons on the wheel control the articulation (note lengths). Presently, the dynamics (loudness) can be coupled with the acceleration or independently engaged (through button control). In the examples in this paper, dynamics are linked directly to the acceleration. In previous work, we have mentioned only in passing the mapping of Brahms' *Hungarian Dance No. 5* to a road in ESP. This paper will systematically describe multiple interpretations of *Dances Nos. 2 and 3* from the same set of pieces, using them as detailed case studies of how an interpretation is created and mapped to a road in the ESP interface.

## 1.2 Related Work

Conductor programs, tools or software that control accurate playback of notes while allowing the user to focus only on features of interpretation (tempo, dynamics, etc.) have been around since the turn of the century (see online review by Malinowski, [10]). Some of the most recent additions to this distinguished line of work is the Air Worm [3], which allows a user to control the tempo and loudness of an audio recording of a piece of music that has been stripped of its expressive nuances, and a literal gestural conducting interface developed in the

Personal Orchestra project [1]. Control of expressive parameters are also part of accompaniment programs (see for example the Music Plus One system described in [12]), in which the soloist acts as the conductor.

The link between locomotion and expressive performance is a well-established one (see for example, [13]). Although expressive performance has primarily been linked to human gait and dance steps [6], anecdotal evidence suggests that performing a piece of music also shares many similarities with the act of driving a car.

This figure shows a musical score for Hungarian Dance No. 2, focusing on symmetric phrasing. The score is presented in a grand staff format (treble and bass clefs) with a 2/4 time signature. It is divided into four systems, with measures 7, 15, 22, and 28 marked at the beginning of each system. The notation includes various musical symbols such as notes, rests, and dynamic markings, illustrating the symmetrical structure of the piece.

Figure 2. *Hungarian Dance No. 2*, symmetric phrasing.

This figure shows the same musical score for Hungarian Dance No. 2, but with gestural annotations. The score is presented in a grand staff format (treble and bass clefs) with a 2/4 time signature. It is divided into four systems, with measures 7, 15, 22, and 28 marked at the beginning of each system. The notation includes various musical symbols such as notes, rests, and dynamic markings. Large, stylized grey arrows and wavy lines are overlaid on the score, indicating specific gestures or movements associated with the music.

Figure 3. *Hungarian Dance No. 2*, gestures.

Entertainment companies focusing on music games have also used the metaphor of movement through space for travel through musical time. See, for example, Harmonixmusic's [8] PS2 games *Amplitude*, *Frequency*, and *Guitar Hero*. Their focus is on generating music as a path that is traversed repeatedly in a loop. Such music games differ from ESP, in which no notes are altered, added or deleted in the process of a performance. Instead, ESP focuses on the synthesis of musical expression. The only parameters under the user's control are those that pertain to the rendering of expression, i.e. tempo, dynamics, and articulation. [2] contains further descriptions of projects and research related to ESP.

## 2. FROM INTERPRETATIONS TO ROADMAPS

In musical interpretation, the choice of structural groupings can exist on multiple levels, and exhibit complex and embedded patterns. Such levels of phrasing and embedded groupings can be made apparent by the manipulation of the tempo and loudness. As Todd [13] discovered, crescendo and decrescendo shapes of varying dynamic ranges indicate the level of embedding in a system of phrases. Furthermore, the dynamics increase and decrease with the tempi.

In ESP, the road guides users to show the grouping structures through a combination of straight and curved sections. For each interpretation, we create a road that encourages the particular set of grouping structures. Bends are inserted at the grouping boundaries to encourage slowdowns of varying degrees. The curvature of a road bend corresponds to the level of embedding in the phrase structure, and the degree of slowdown desired. Because the dynamics are coupled with the acceleration in ESP, the dynamics do increase and decrease with the tempi.

The mapping from an interpretation to a road follows a few basic principles: (1) adjacent phrases at the lower levels (smaller groupings) are delineated by a bend in the same direction as the general curvature of the road; (2) larger groupings at higher levels are delineated by curve that marks a change in direction; and, (3) larger boundaries, such as those between sections, have steeper bends than elsewhere.

The positions of the curves are adjusted to account for variability in response time to road bends. Human response time to the bends vary with the degree of curvature. In driving, the user tends to decelerate before entering a bend. The distance before the bend at which this deceleration first occurs depends on the sharpness of the curve. Gentler bends require less slowdown, and hence the deceleration occurs closer to the curve. Sharp bends demand more care in driving, and encourage slowdowns well before the curve.

In the following sections, we show examples of how different interpretations of Brahms' *Hungarian Dances Nos. 2 and 3* are mapped to different roads in ESP. We choose Brahms' *Hungarian Dances* as example pieces because of the extreme ranges of tempi and dynamics (as well as articulation) encapsulated within each piece in the set. They form an excellent test set for the ESP interface.

### 2.1 Mapping Study I: Roadmaps for Brahms' *Hungarian Dance No. 2*

In this section, we present two expressive interpretations of the first 32 bars of Brahms' *Hungarian Dance No. 2* and their corresponding roadmaps. With the two repeats, we have a total of 64 bars of music in an AABB form. The first interpretation is symmetric and square, grouping bars by powers of two. The symmetric grouping at different hierarchical levels, indicated by phrase arcs, is shown in Figure 2. The bar groupings can be represented numerically (by bar numbers) as:

$$[(1,4)(5,8)] \{ [(9,10)(11,12)] [(13,14)(15,16)] \} \dots$$

The second interpretation, shown in Figure 3, stresses melodic peaks (shown by single grey-lined mounds in bars 3 and 7) and areas of harmonic tension (indicated by squiggles) and release (forward arrows), which lead to musical gestures that often do not correspond to the symmetric phrase structures.

The resulting roads corresponding to the symmetric phrasing and gesture-based interpretations are shown in Figures 4 and 5 respectively. The maps on the green backgrounds are the birds' eye views provided to the user; the color-segmented maps on the right show the AABB sections of the excerpt mapped, indicated by red, orange, green, and blue segments respectively.

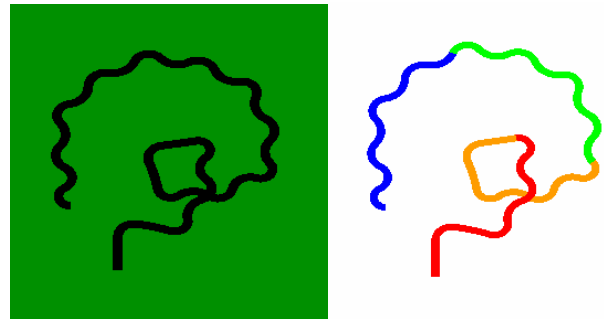


Figure 4. *Hungarian Dance No. 2* roadmap, symmetric.

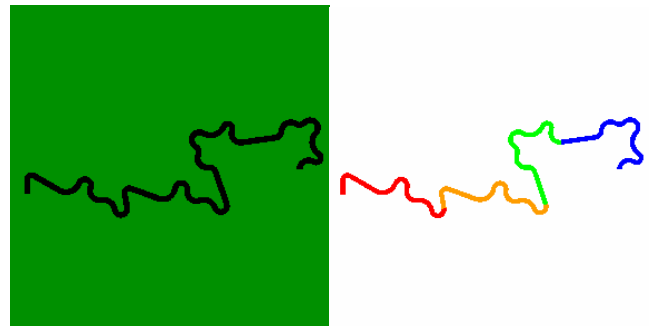


Figure 5. *Hungarian Dance No. 2* roadmap, gestures.

Table 1 shows an example of a curvature table from which the roadmap in Figure 4 is constructed. In this example, we chose to alternate the direction of the curves between left (L) and right (R). Phrase boundaries at the lowest level are marked by small curves, and higher level phrase boundaries correspond to larger bends.

**Table 1. Curvature data for constructing roadmap for Brahms' *Hungarian Dance No. 2*, symmetric phrasing**

Curve Number	Length (beats)	Degree Turn	Direction
1	5.5	1	L
2	3.5	100	R
3	4.5	10	L
4	3.5	120	R
5	4	80	L
6	4	120	R
7	4	100	L
8	4	150	R
9	4.5	1	L
10	3.5	100	R
11	4.5	10	L
12	3.5	120	R
13	4	80	L
14	4	120	R
15	4	100	L
16	4	150	R
17	4	80	L
18	4	100	R
19	4	80	L
20	4	120	R
21	4	80	L
22	4	100	R
23	4	80	L
24	4	120	R
25	4	80	L
26	4	100	R
27	4	80	L
28	4	120	R
29	4	80	L
30	4	100	R
31	4	120	L
32	4	150	R

## 2.2 Mapping Study II: Roadmaps for Brahms' *Hungarian Dance No. 3*

In this section, we propose two grouping solutions for the first 36 bars of Brahms' *Hungarian Dance No. 3*, shown in Figures 6 and 7. The third Hungarian Dance is marked by an atypical 3-bar phrase structure, which gives rise to asymmetries and many possibilities for alternate groupings. The first 24 bars (counting the repeated bars) are shown in Figures 6 and 7. Bars 25-36 (not shown) repeat material in bars 1-12, resulting in an ABA structure.

In the first interpretation, we take the most obvious (naïve) phrase grouping arising from strict adherence to the bar-lines (see Figure 1). The lowest level grouping follows the barline, one group per bar. This results in a march-like two-step rhythm in each bar in Section A (the first 12 bars, counting repeated bars). Each set of three bars form the next level grouping, and are shown by larger phrase arcs in Figure 6. In Section A, there is the option to create a middle level grouping of 1:2 or 2:1 within each three-bar set. We have chosen the 2:1 mid-level grouping for Section A. In the B section, again, the lowest level grouping resides at the bar, and the high level grouping consists of three bars each. Due to the continuous flow of this new material, we do not introduce a mid-level grouping here.

**Figure 6. *Hungarian Dance No. 3*, naïve interpretation.**

**Figure 7. *Hungarian Dance No. 3*, nuanced interpretation.**

The resulting roadmap for this first (naïve) interpretation is shown in Figure 8. As before, the map on the green background shows



the user's bird's eye view, while the color-segmented map indicates the ABA segments (red, green, and red again). The red segments are identical, and cross each other in the two-dimensional map. In the color-coded map, the yellow header contains the upbeat note and marks the beginning of the piece.

Due to the brevity of the lowest level phrase in this piece, the junction between each phrase pair at the bar level is marked by small curves in the same direction as the preceding segment. If this junction corresponds to a phrase boundary at a higher level, then the road changes direction. The B section is constructed in a similar fashion. Between the A and B sections, we increased the curvature of the bend to mark the large structural boundary. To prevent too much of a slowdown (resembling stalling) in the second set of three bars, the curve here is made more gentle to encourage forward motion.

In a second, more nuanced, interpretation (shown in Figure 7), we choose structural groupings that do not adhere to the barlines; instead, they follow the melodic grouping principles of longer notes and larger interval leaps marking boundaries, short upbeats being grouped with the longer notes, and of repeated patterns and symmetries signaling groups. For example, the opening upward gesture (C, CDE, F) is followed by another upward gesture (F, Bb, D, C) in the right hand, forming the first two low level groups. At the next level up, the second set of four eighth notes follows a pattern similar to the first set of four eighth notes, forming two higher level groups. As in the previous interpretation, the larger phrases consist of three bars each, with the option to subdivide into 1:2 or 2:1, the difference being that the groupings in this section are not aligned with the barlines. For a change, we have chosen the 1:2 mid-level grouping in Section A.

In Section B, again the phrases are offset from the barlines by an eighth note. In this more nuanced interpretation, we add a mid-level grouping that subdivides the high level three-bar phrase into two parts, each lasting six eighth notes, cutting across the barlines, and resulting in a pleasing symmetry.

This more nuanced interpretation encouraged us to examine variance in grouping significance for structural boundaries at the same level. We found that not all boundaries at the same level should be treated the same way. For example, even though the boundaries between the four sets of three bars in section B (green)

were at the same level, we chose to make the middle one less than the others so as to continue the forward motion in the music.

The resulting roadmap for this second (nuanced) interpretation is shown in Figure 9. As for the naïve interpretation map, the lowest level phrase boundaries were marked by same-direction bends; the road direction changes for higher-level boundaries. Because this interpretation does not follow the bar lines, the distances of the phrase groupings are not as symmetrical as that for the naïve interpretation. We also found that because of the range of curvature types and proximity of curved sections, we had to be more careful with planning the placement of the curves to adjust for different deceleration patterns when approaching a curve.

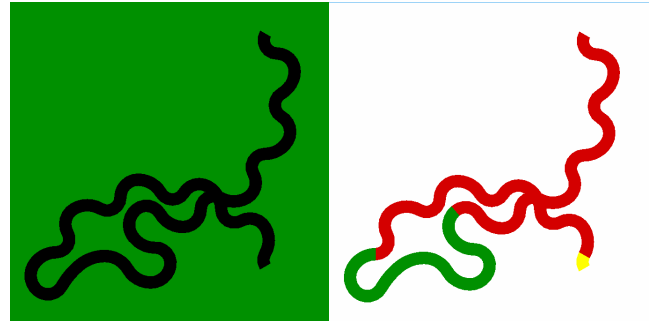


Figure 8. *Hungarian Dance No. 3* roadmap, naïve interpretation.

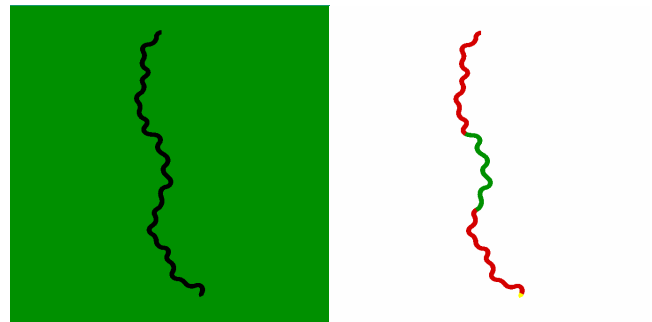


Figure 9. *Hungarian Dance No. 3* roadmap, nuanced interpretation.

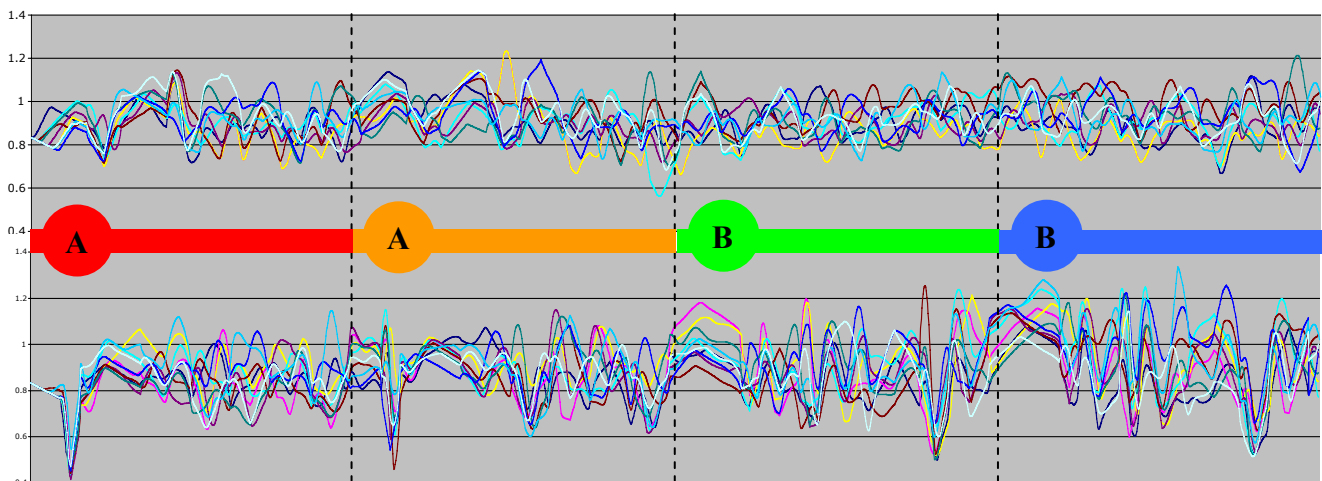


Figure 10. Tempo curves for *Hungarian Dance No.2* from (top) symmetric phrasing and (bottom) gesture-based maps.

### 3. FROM ROADMAPS TO PERFORMANCES

In this section, we present and analyze the driving (performance) patterns resulting from the maps of the different interpretations of the two Hungarian Dances. One of the authors (Jie Liu), an experienced user of the ESP system, created ten performances each on every one of the roads described in Section 2. The driving style was careful, slowing down and bends as necessary in order to minimized collisions with the road edge. The numerical data sets are presented in this section; the MIDI files of the performances can be obtained from [www-rcf.usc.edu/~mucoaco/ESP/2006-AMCMM](http://www-rcf.usc.edu/~mucoaco/ESP/2006-AMCMM) for validation. The effect of an expressive performance cannot be fully comprehended by simply examining tempo graphs. Thus, the reader is strongly encouraged to listen to the examples posted by visiting the url given above.

#### 3.1 Performance Results I: Brahms' Hungarian Dance No. 2

The resulting tempo curves for the two interpretation maps, symmetric phrasing and gesture-based, for Brahms' *Hungarian Dance No. 2* are shown in Figure 10. The corresponding MIDI files are named *Brahms\_HD2\_symmphrasing\_m.mid* and *Brahms\_HD2\_gesture\_n.mid* respectively in the MIDI repository, where  $m, n = 1, \dots, 10$ . In Figure 9, Each line in the two graphs shows the result of one drive through a path – there are ten in total for each map, all produced by the same driver (Jie Liu). The ten iterations through the same path show the patterns of tempo variance over different iterations over the same road. For this data set, the driver took a break after each set of five performances. On the graphs, each section of the map is indicated by A or B labels, with color-codes that parallel the road segments. The tempi are given in beats per second; for example, a tempo level of 1 corresponds to 60 beats per minute.

In general, the gesture-based roadmap resulted in more extreme speedups and slowdowns, as intended. In other words, the gesture-based map resulted in larger musical gestures. For example, each section A is marked by a steep slowdown near its beginning, a gesture at the right place, but of magnitude perhaps

larger than intended; and each section B begins with an acceleration in tempo, as intended in the road design. The musical effect can best be examined through hearing playbacks of the MIDI files recorded from these performances.

The systematic variations in the symmetric phrasing map are less easy to define by inspection. Recall that even though bends encourage slowdowns, the user ultimately still has control over the expressive gestures through their dexterity at the wheel and real-time auditory feedback of the unfolding piece. The latitude afforded the user in making performance decisions makes it difficult to map every slowdown to a particular bend.

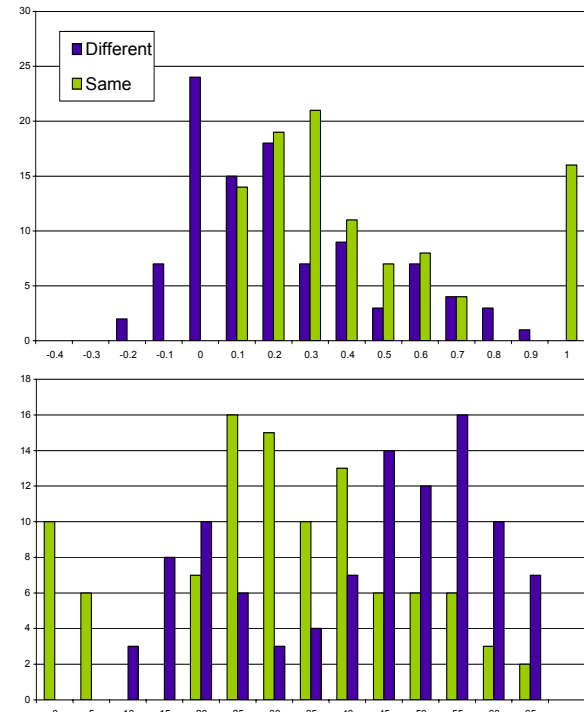


Figure 11. *Hungarian Dance No.2*: histograms of (top) correlation coefficient and (bottom) SSE values for performances from same and different roads.

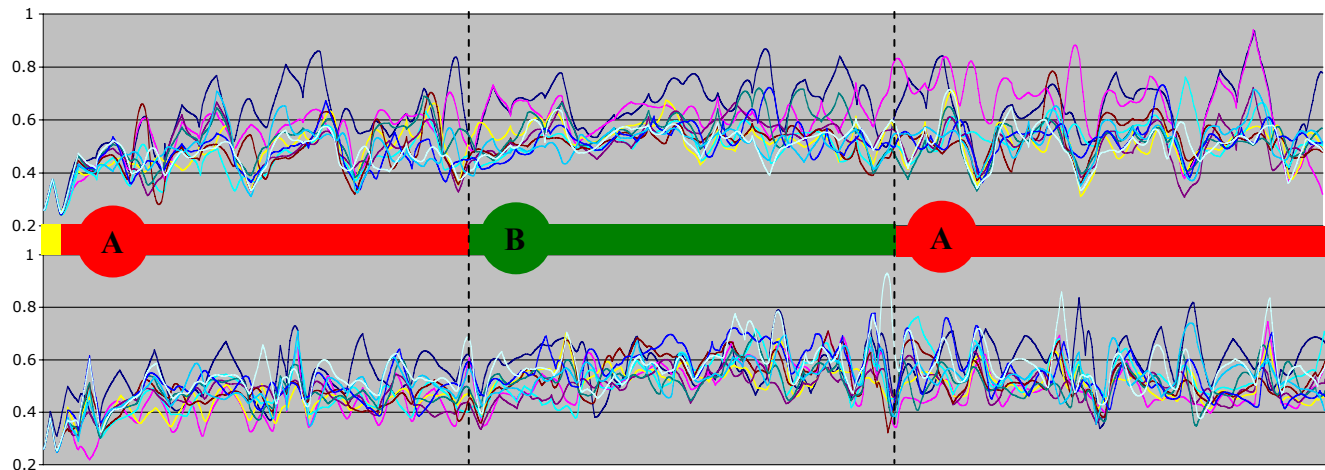


Figure 12. Tempo curves for *Hungarian Dance No.3* from (top) naïve and (bottom) nuanced interpretation maps.



To demonstrate empirically that the two sets of tempo time series are significantly different, we calculated, for each pair of performances, the correlation coefficient and the sum of squared errors (SSE) between their tempo curves. The equations for the correlation coefficient and the SSE measures for two tempo sequences,  $\{x_i\}$  and  $\{y_i\}$ , are as follows:

$$r = \frac{1}{n-1} \sum_i \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right),$$

$$\text{and SSE} = \sum_i (x_i - y_i)^2,$$

where  $\bar{x}$  and  $\bar{y}$  are the means, and  $s_x$  and  $s_y$  are the standard deviations of  $\{x_i\}$  and  $\{y_i\}$  respectively

The distributions of these values are shown in Figure 11. The correlation coefficient values for same road and different road performances come from distributions with different means, 0.16 for performances from different roadmaps versus 0.39 for performances from the same roadmap. The SSE values ranged from 0 to 65. The SSE distributions for same and different road performances are each bimodal with significantly different peaks, at 20 and 55 for performances from different roadmaps versus 0 and 25 for performances from the same roadmap. The mean SSE for different road performances is 39.2, while that for same road performances is 28.4. Both the correlation coefficient and the SSE measures show that, in general, performances resulting from the same road are more similar to each other than those from different roads.

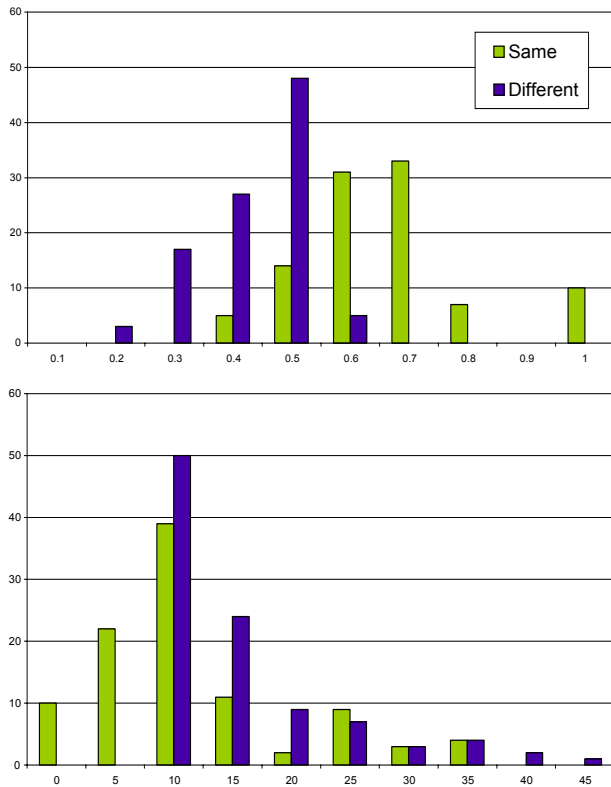


Figure 13. *Hungarian Dance No.3*: histograms of (top) correlation coefficient and (bottom) SSE values for performances from same and different roads.

### 3.2 Performance Results II: Brahms’ *Hungarian Dance No. 3*

The performance (driving) results (tempo graphs) for the naïve and nuanced interpretations of Brahms’ *Hungarian Dance No. 3* are shown in the two graphs in Figure 12. The corresponding MIDI files are named *Brahms\_HD3\_naive\_m.mid* and *Brahms\_HD3\_nuanced\_n.mid* respectively in the MIDI repository, where  $m, n = 1, \dots, 10$ . As before, the same driver (Jie Liu) drove on each path ten times to create the twenty performances shown on the two graphs.

Again, by inspection, the two different roads produced different tempo trajectory patterns. One prominent difference is the marked tempo slowdown exhibited by all performances on the nuanced interpretation road in the return from the B to A section, as shown by the local minima of the entire bunch of tempo curves around the second dotted vertical lines in Figure 12.

As before, we show the distributions of the correlation coefficients and the SSE’s of performances from the same and different roads in Figure 13. The correlation coefficients of same and different road performances come from distributions with different means. The mean correlation coefficient value for same road performances is 0.61, while that for different road performances is 0.38. The SSE values ranged from 0 to 41. Although the SSE peaks for same and different road performances are not visibly different, their underlying means are 9.4 and 13.0, respectively, for same and different road performances. Again, both measures show that, overall, performances resulting from the same road patterns are more similar to one another than those from different roads.

## 4. CONCLUSIONS AND DISCUSSION

We have demonstrated, in this paper, examples of multiple interpretations of the same scores, and how this could map to different roads in the ESP interface. We have presented empirical results of driving/performance outcomes using these roads in ESP, and analyzed their tempo time series. The general trends of slowdowns near curvature extremities (which correspond to structural boundaries or gestural slowdowns) and accelerations when the road is straight can be observed in many cases. Statistics show that the road has the effect of constraining the performances to a particular style, with specific gestures, although the user (driver) still maintains some latitude in expressive control. Overall, the interpretation maps for the two Hungarian Dances resulted in expressively pleasing performances.

A natural question arises on the role of auditory feedback in the performance/driving of a piece. For an expert driver, the speeds will likely be much higher, with greater variance, without the moderating effects of auditory feedback. Thus, the tempo trajectory recorded in each performance is not only the result of the road mapped according to an interpretation of the score, it is also guided by the user’s hearing of the piece. Further experiments will have to be designed to determine the extent of the effect of auditory feedback on expressive control.

Although many curvature maxima and minima, which correspond to right or left turns, induced tempo slowdowns, the actual local tempo minima are sometimes a little before, sometimes a little after, the extreme points. Occasionally, little slowdown, or even speeding up, is observed at a bend. We observed that the link between

deceleration and curvature extremity is more pronounced in the cases when the curves can be more reliably predicted through adequate lead-ins or periodic spacing. When there is little time between curves, their slowdown effect is reduced, and the performance will show low correlation between tempo slowdowns and curvature peaks.

Aural inspection (listening) shows that the performances generated on the roadmaps constructed from particular interpretations of the score did indeed achieve the desired effects marked in the interpreted scores. The sense that the guided performance worked as planned is not always readily captured in the tempo data (which is correlated with the loudness because of the link between loudness and acceleration). For example, a positive tempo change of 0.2 beats per second, is sometimes smoothed over by the human ear and regarded as stable, even though visually the graph shows a significant tempo increase. The gap between perception and signal leads us to postulate that perhaps the instantaneous tempo should be considered a fuzzy quantity in the extraction and understanding of musical gesture.

More work remains to be done to devise new and more precise ways of evaluating the effectiveness, and degree of influence, of the ESP interface. The challenges include the subjectivity of music performance evaluation, and the fact that unique and outlier performances can often be desirable ones. Systematic evaluation in the future will likely include investigations such as (1) the impact of a road pattern on driving (performance) behavior; (2) the impact of auditory feedback on driving (performance) choices; and, (3) the effect of repetition on expressive choices over the same material. It would be desirable to validate the experimental results across multiple users.

One way that music creates interest is by generating expectations and meeting them through repetition. For example, in Brahms' *Hungarian Dance No. 2* example, the structure is AABB, where the A and B sections are each repeated; in the *Hungarian Dance No. 3* example, the A section returns at the end of the ABA structure. A real performer will likely play the second time a section appears a little differently than the first. The same is hypothesized to be true in performances generated through the ESP interface. Evidence for this hypothesis can be found, for example, in the two iterations each of A and B sections in performances of Brahms' *Hungarian Dance No. 2* produced from the gesture-based map (shown in the lower half of Figure 10). Future experiments will focus on the intentional expressive choices exercised by the user when a section is repeated. Such investigations will also reveal the nature of the effect of listening on performance choices.

## 5. ACKNOWLEDGMENTS

We thank Dennis Leung for his help in encoding the Brahms *Hungarian Dances* in Sibelius format for generating expressionless MIDI and score information. This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 0347988. The work made use of the Integrated Media Systems Center Shared Facilities supported by NSF under Cooperative Agreement No. EEC-9529152. Any opinions, findings, conclusions,

or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

## REFERENCES

- [1] Borchers, J., Hadjakos, A., and Mühlhäuser, M. MICON: A Music Stand for Interactive Conducting. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME'06)* (Paris, France, June 4-8, 2006), 254-259.
- [2] Chew, E., François, A. R. J., Liu, J., and Yang, A. ESP: A Driving Interface for Expression Synthesis. In *Proceedings of the International Conference on New Instruments for Musical Expression (NIME'05)* (Vancouver, B.C., Canada, May 25-28, 2005), 224-227.
- [3] Dixon, S., Goebel, W., and Widmer, G. The PERFORMANCE WORM: Real-Time Visualization of Expression based on Langner's Tempo-Loudness Animation. In *Proceedings of the International Computer Music Conf. (ICMC 2002)* (Göteborg, Sweden, September 16-21, 2002).
- [4] ESP – Expression Synthesis Project. [www-rcf.usc.edu/~mucoaco/ESP](http://www-rcf.usc.edu/~mucoaco/ESP)
- [5] François, A. R. J., and Chew, E. An Architectural Framework for Interactive Music Systems. In *Proceedings of the International Conference on New Instruments for Musical Expression (NIME'06)* (Paris, France, June 4-8, 2006), 150-155
- [6] Friberg, A., Sundberg, J., and Fryden, L. Music from motion: Sound level envelopes of tones expressing human locomotion. *Journal of New Music Research*, 29:3 (2000), 199-210.
- [7] Gabrielsson, A. Music Performance. In D. Deutsch (Ed.): *The Psychology of Music*, Vol. 2, pp.501-602. San Diego, CA: Academic Press, 1999.
- [8] Harmonixmusic. [www.harmonixmusic.com](http://www.harmonixmusic.com)
- [9] Liu, J., Chew, E., and François, A. R. J. User Interface Design in ESP: Ensuring Tempo Smoothness in the Mapping of Driving to Music Performances. In *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technologies (ACE 2006)* (Hollywood, CA, USA, June 14-16, 2006).
- [10] Malinowski, S. The Conductor Program [www.musanim.com/tapper](http://www.musanim.com/tapper)
- [11] Palmer, C. Music Performance. *Annual Review of Psychology*, 48 (1997), 115-38.
- [12] Raphael, C. Orchestral musical accompaniment from synthesized audio. In *Proceedings of the International Computer Music Conf. (ICMC 2003)* (Singapore, Singapore, September 29 - October 4, 2003).
- [13] Todd, N. P. M. The kinematics of musical expression. *J. of the Acoustical Society of America*, 97, 3 (March 1995), pp.1940-1949.