Switching State Space Models for Interpreting Musical Dynamics

Robert Granger

Department of Statistics, Indiana University

August 22, 2020

Abstract

In this paper, we attempt to illuminate a performer's musical intentions using a Markov-switching state space model. Specifically, the model we propose is designed for the dynamics of classical music. Unlike many other models of trend estimation, this model requires the estimation of a series of parameters. These parameters can be viewed as hidden musical attributes of a performance that would not be revealed through the calculation of simple statistics. For this analysis, we use dynamics data of 46 different performances of Chopin's Mazurka Op. 68 No. 3. from the Centre for the History and Analysis of Recorded Music (CHARM) Mazurka Project.

Keywords: classical music, hidden Markov model

1   Introduction

In this paper, we attempt to illuminate a performer's musical intentions surrounding the dynamics of classical music. Through the use of a Markov-switching state space model, the series of dynamics values within a single performance are carefully smoothed with the goal of differentiating the perceived intentions of the performer from the observed dynamics found in the data. Classical music is different from other types of music as a composer creates a piece, but many performers may play it adding their own interpretation. While, the composer gives some specific direction like the pitch of notes to play, the instructions regarding dynamics are relatively vague. For example, when performing a piece, a performer will observe a p which stands for *piano*, indicating this section should be played "softly". Meanwhile, an f stands for *forte* and indicates the given section should be played "loudly". While we would expect the *forte* section to be played louder than the *piano* section, no direction is given on how softly or loudly the specified section should be played. The performer may also attempt to ease into a *piano* section from a *forte* section or they may want to make a sudden change. These decisions are left to the interpretation of the performer.

There are a variety of different trend estimating techniques, but we propose using a Markov-switching state space model for two reasons. First, the technique should incorporate actions found within a performance. Often times, performers make dramatic changes in dynamics to start a new section, something we want our model to identify, not smooth out as noise. Likewise, this may extend to individual notes where intentional emphasis is used by the performer. Using smoothing techniques like a moving average or spline would ignore these key elements. The second reason for using this type of model is that it requires the estimation of certain parameters. These parameters offer additional insight and can be viewed as hidden attributes of a performance. Attributes found through simple calculation of averages and variances would ignore the complexity of the decisions made during a performance.

Similar analysis of modelling musical performances with Markov-switching state space models has been performed on tempo data (McDonald et al. 2019, Gu & Raphael 2012). When modeling tempo, the expectation is for the performer to hold a constant tempo, with

some periods of potential speeding up or slowing down. This same expectation cannot be said for the dynamics and thus requires a different model setup. Although a Markov-switching state space model can still be used, the difference comes in the selection of different discrete states and parameter matrices. While Gu and Raphael did give some attention to modeling dynamics, this paper looks to expand their work.

The paper procedes as follows: Section 2 begins with a general description of Markov-switching state space models followed with specific details regarding the states and the parameter matrices used to model musical dynamics. Section 3 explains the algorithm used to evaluate the model and discusses results for several performances. Section 4 briefly concludes with potential uses of this model along with potential future areas of research.

2   The Markov-Switching State Space Model

State space models are commonly used to model time series observations in the presence of hidden, continuous states. As a result of the state space framework, the observations are viewed as independent conditional on the hidden states whereas these hidden states will follow a vector autoregressive process. Adding assumptions of linearity and normal error produces what is commonly referred to as the general linear Gaussian state space model (Durbin & Koopman 2012). It takes the form

$$
\begin{aligned}
y_t &= C_t + D_t x_t + \epsilon_t, \quad \epsilon_t \sim N(0, G_t) \\
x_{t+1} &= A_t + B_t x_t + \eta_t, \quad \eta_t \sim N(0, H_t), \quad x_1 \sim N(x_0, P_0)
\end{aligned}
\tag{1}
$$

where the first part of Equation 1 is known as the observation equation and the second part is known as the state equation. The vector y_t consists of the known observations at each time period, t, whereas the vector x_t consists of the unobserved, continuous states on which y_t is dependent. The observation error, ϵ_t, and the state equation error, η_t, are assumed to be serially independent and independent of each other.

In the typical state space framework, the matrices A_t, B_t, C_t, D_t, G_t, and H_t are allowed to vary across time but are known. If there are a finite number of perceived structures for these matrices, and the given structure is unknown at time, t, a Markov-switching state space model can be used. This model assumes there are some underlying discrete states,

s_t, that transition over time through a Markov process. Making this slight adjustment to Equation 1 yields the following model which will be used as a basic framework for modeling music dynamics.

$$y_t = C_t(s_t) + D_t(s_t)x_t + \epsilon_t, \quad \epsilon_t \sim N(0, G_t)$$
$$x_{t+1} = A_t(s_t) + B_t(s_t)x_t + \eta_t, \quad \eta_t \sim N(0, H_t), \quad x_1 \sim N(x_0, P_0) \tag{2}$$

When it comes to musical dynamics, rarely do musicians attempt to play with the same dynamics or loudness throughout the entirety of a piece. Most of the time we expect the musician to steadily change the loudness from note to note; however, there may exist moments where the musician deviates greatly from the trend with either louder or softer than usual notes. A single note may be played more loudly because the performer is specifically trying to add emphasis. Adding this emphasis may be on the performer's own perogative or may be dictated by the piece with an accent, ">". On the other hand, a single note appearing in the data more softly is likely not part of the performer's intent. If a note in the data is played more softly, it may be because the performer missed a note or did not hit it as hard as intended. It could also be something went wrong in the data creation step. Regardless, we do not attempt to identify why the dynamics data indicates an unusually soft note, but it is important to include this in our model for statistical estimation purposes. We tried estimating without explicitly incorporating these low sounds in the model, but the excessive statistical noise caused the model to falsely indicate starts of new smooth progressions. Therefore, in order to model the described behavior, we propose the following four discrete states:

s^1: The musician selects a new value for loudness.

s^2: The musician continues the dynamics in a steady way.

s^3: The musician plays a single note more loudly.

s^4: The musician plays a single note more softly.

The observation, y_t is the univariate loudness of the note at each time period, t. In order to allow the dynamics to progress steadily, the continuous hidden states, x_t, follow a

Table 1: Parameter matrices for the switching state space model.

| States	Parameter Matrices					
S	A	B	C	D	G	H
s^1	$\begin{pmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	0	$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$	σ_ϵ^2	$\begin{pmatrix} \sigma_0^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_2^2 \end{pmatrix}$
s^2	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$	0	$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$	σ_ϵ^2	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
s^3	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$	μ_c	$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$	σ_ϵ^2	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
s^4	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$	μ_e	$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$	σ_ϵ^2	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

process that allows for piece-wise quadratics. At each time period, x_t, is a vector of length three,

$$x'_t = (x_t^0, x_t^1, x_t^2),$$

where x_t^0 is the loudness, x_t^1 is the first order difference, and x_t^2 is the second order difference. The aim is to maintain this smooth progression even when the musician plays a single note more loudly or softly. The states s^3 and s^4 are therefore implemented by adding a constant in the observation equation as opposed to changing the state equation. The model is similar to that proposed in Gu & Raphael (2012), but extended to include these additional states. Table 1 shows the parameter matrices for the four states.

The final part of the switching state space model is designing the Markov process for

Figure 1: Transition diagram.
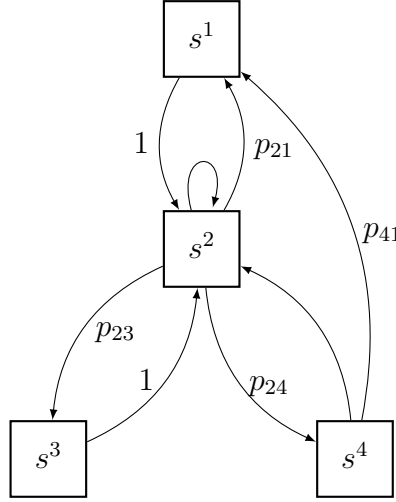
transitioning between the discrete states which is displayed in Figure 1. The first state, s^1, allows for the selection of a new loudness which then transitions into the smooth progresson state, s^2, with probability 1. When arriving in s^2, the next time period's discrete state can be any of the possible states including itself. If in this progression, we move to a sudden loud note, s^3, or a sudden soft note, s^4, then we have the opportunity to continue the smooth progression, s^2, or start a new smooth progression, s^1. It is not permissible though to immediately then play another sudden loud or soft note.

In total, there are 13 unknown parameters, $\theta \in \Theta$, spread across 4 different discrete states, $s^i \in S$, through T time periods. A summary of the parameters to be estimated are: $\Theta = \{\mu_c, \mu_e, \sigma_\epsilon^2, \mu_0, \mu_1, \mu_2, \sigma_0^2, \sigma_1^2, \sigma_2^2, p_{21}, p_{23}, p_{24}, p_{41}\}$. Note this vector contains only four probabilities. With four states, the transition matrix could potentially have up to twelve probabilities requiring estimation; however, because of the restrictions placed on the possible transitions, this number is reduced to four. In the next section, we discuss finding smoothed estimates of the dynamics which requires estimation of these parameters along with the continuous and discrete states.

3   Evaluating the Model

Evaluation of the model requires the estimation of the parameters, $\theta \in \Theta$; the discrete states, $\{s_t\}_{t=1}^T$; and the continuous hidden states $\{x_t\}_{t=1}^T$. These estimates are obtained using only the assumptions and structure of the model along with the observed dynamics, $\{y_t\}_{t=1}^T$. Once these estimates are obtained, we then compute predicted values, $E[y_t|...]$, which can be thought of as the performer's intended dynamics. Hence, the aim in estimating the intended dynamics is attempting to remove ϵ_t, which can be seen as unintended deviations from the desired loudness.

3.1   The Algorithm

When the parameter values, $\theta_i \in \Theta$, and the discrete states, $\{s_t\}_{t=1}^T$, are known, the Kalman filter (Kalman 1960) can be used to find estimates of the continuous hidden states, $\{x_t\}_{t=1}^T$, along with the likelihood. While the Kalman filter provides an easy way to compute the likelihood, the estimate of x_t is obtained using only observations coming before time t. In order to use all information, we implement the Kalman smoothing algorithm introduced in (Rauch et al. 1965). This algorithm provides an estimate, $\hat{x}_t = E[x_t|y_1, ...y_T]$, which can be used to compute a fitted or "smoothed" value, $\hat{y}_t$.

The Kalman smoother provides a closed form solution to obtaining the maximum likelihood estimates of our continuous hidden states, but what if the discrete states, $\{s_t\}_{t=1}^T$, are unknown? We can obtain the most likely set of discrete states by simply running the Kalman smoother algorithm on every possible combination of discrete states and choosing the set of discrete states that yields the largest likelihood. This may work if the number of discrete states and/or time periods is small; however, the number of state combinations to check may be large and can be computed as $|\{s^i \in S\}|^T$. For example, the music dynamics model presented in the previous section has four discrete states and the piece to be evaluated, Chopin's Mazurka Op. 68 No. 3, has 231 notes. This brings the grand total of state combinations to check to $4^{231} \approx 1.19 \times 10^{139}$. Of course, many of these state combinations could be removed due to 0 likelihood given the restrictions on the state transitions, but even with this taken into account, the number of state combinations is far too large. To overcome this issue, we use the Discrete Particle Filter as described in McDonald et al.
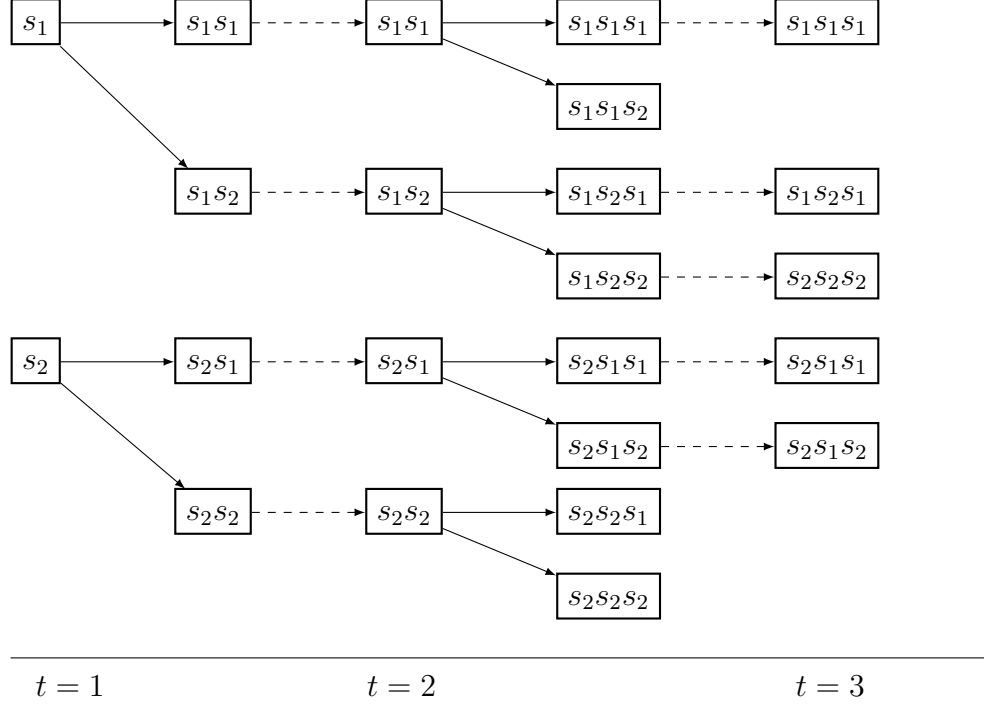
Figure 2: Discrete Particle Filter with 2 discrete states and capping the maximum number of stored particles at 5 for each time iteration.

(2019). We begin by estimating the partial likelihood iteratively through time at each of the possible discrete state paths using the Kalman Filter. Each of these paths is known as a particle with only the previous states and partial likelihood being saved. This process would still require checking all $|\{s^i \in S\}|^T$ paths, so to get around this, we implement the greedy search algorithm called Beam Search. This search algorithm requires selecting a maximum number of particles to save at each time iteration and discards the rest.

Figure 2 illustrates this concept with 2 discrete states and the maximum number of particles set at 5. There are four possible state paths to check when passing from $t = 1$ to $t = 2$ which is less than the maximum particle number so all four paths are saved. When moving from $t = 2$ to $t = 3$, the number of paths increases to 8. Since only 5 paths are allowed, three of these paths must be dropped and will no longer be considered as possible solutions. In order to decide which paths are to be saved a sampling procedure must be performed. One sampling procedure proposed by (Tugnait 1982) is to simply keep the paths that have the highest likelihood. Another sampling procedure proposed by (Akashi & Kumamoto 1977) is to randomly sample one particle out of the total number of states

---
Algorithm 1: Solving Music Dynamics Model

    Input Y, a vector of observed dynamics;

    Initialize Θ;

    while *(Stopping Criteria)* **do**

        Create Matrices A, B, C, D, G, H at each t for each $s^i \in S$;

        Implement Discrete Particle Filter;

        Compute the log-likelihood: $l(Y|\Theta, S) + l(S|\Theta) + l(\Theta)$;

        Update Θ via Nelder-Mead Optimization or SANN;

    end

---

descended from each of the $t - 1$ particles in proportion to their respective likelihoods. (Fearnhead & Clifford 2003) proposes another stochastic approach but one that minimizes the expected mean squared error between the approximated distribution and the true distribution's probabilities. This approach determines a threshold value such that all particles with likelihood above this value are kept and the remaining particles to be kept are chosen at random with probability equal to their respective likelihoods. Once a sampling technique is selected and the procedure is implemented through all time points, a sample of discrete state sequences is saved the size of the beam width. Since the goal of this paper is to find the most likely sequence of discrete states, the one with the highest likelihood is selected.

The final step is estimating the model parameters, Θ, that maximize the likelihood. This is a difficult task as this model presents a couple of problems: 1) many of the parameters are constrained as variances need to be positive and the probabilities of leaving the same state should sum to 1 while also being constrained between 0 and 1; 2) there are many local maxima making finding a global maxima difficult. To help alleviate these concerns, we can include Bayesian priors on our parameters. Using carefully selected priors allows us to steer the algorithm away from impossible solutions and direct it toward more desired solutions. These issues also lead us to carefully consider our optimization technique. We will use two different methods. The first is the Nelder-Mead algorithm which can find local maxima for unconstrained multidimensional problems without the need for derivative computation (Nelder & Mead 1965). The second is the Simulated Annealing (SANN) algorithm which is

Table 2: Informative prior distributions for Chopin's Mazurka Op. 68 No. 3

Parameter		Distribution	Prior Mean
μ_c	$\sim$	Gamma(100, 0.1)	10
μ_e	$\sim$	-Gamma(100, 0.1)	-10
σ_ϵ^2	$\sim$	Gamma(10, 0.5)	5
μ_0	$\sim$	Normal($\overline{Y}$, 10)	$\overline{Y}$
μ_1	$\sim$	Normal(0, 0.25)	0
μ_2	$\sim$	Normal(0, 0.25)	0
σ_0^2	$\sim$	Gamma(10, 1)	10
σ_1^2	$\sim$	Gamma(3, 1)	3
σ_2^2	$\sim$	Gamma(3, 1)	3
$p_{2,\cdot}$	$\sim$	Dirichlet(5, 85, 5, 5)	0.05, 0.85, 0.05, 0.05
$p_{4,\cdot}$	$\sim$	Beta(2, 8)	0.20

commonly used when dealing with problems with many local maxima in order to estimate the global maxima. Neither of these methods are perfect at avoiding a local maximum and hence the selection of initial values plays an important role. To overcome this challenge, we will select a variety of different starting points drawn randomly from the Bayesian priors.

Algorithm 1 displays the complete procedure. The solution to this model was found using software R (R Core Team 2019). The Nelder-Mead and SANN methods are performed using the package `optimr` by Nash (2019). The discrete particle filter was implemented by extending the package `dpf` by McDonald (2020).

3.2 Results for Chopin's Mazurka Op. 68 No. 3

The model was estimated using observed dynamics from forty-six different performances of Chopin's Mazurka Op. 68 No. 3. This data was reverse conducted (created) by Craig Sapp using Andrew Earis's Expression Algorithm software. The software and data can be found on the website for the Centre for the History and Analysis of Recorded Music (CHARM) Mazurka Project (CHARM 2009). The technique used to create the data is described in
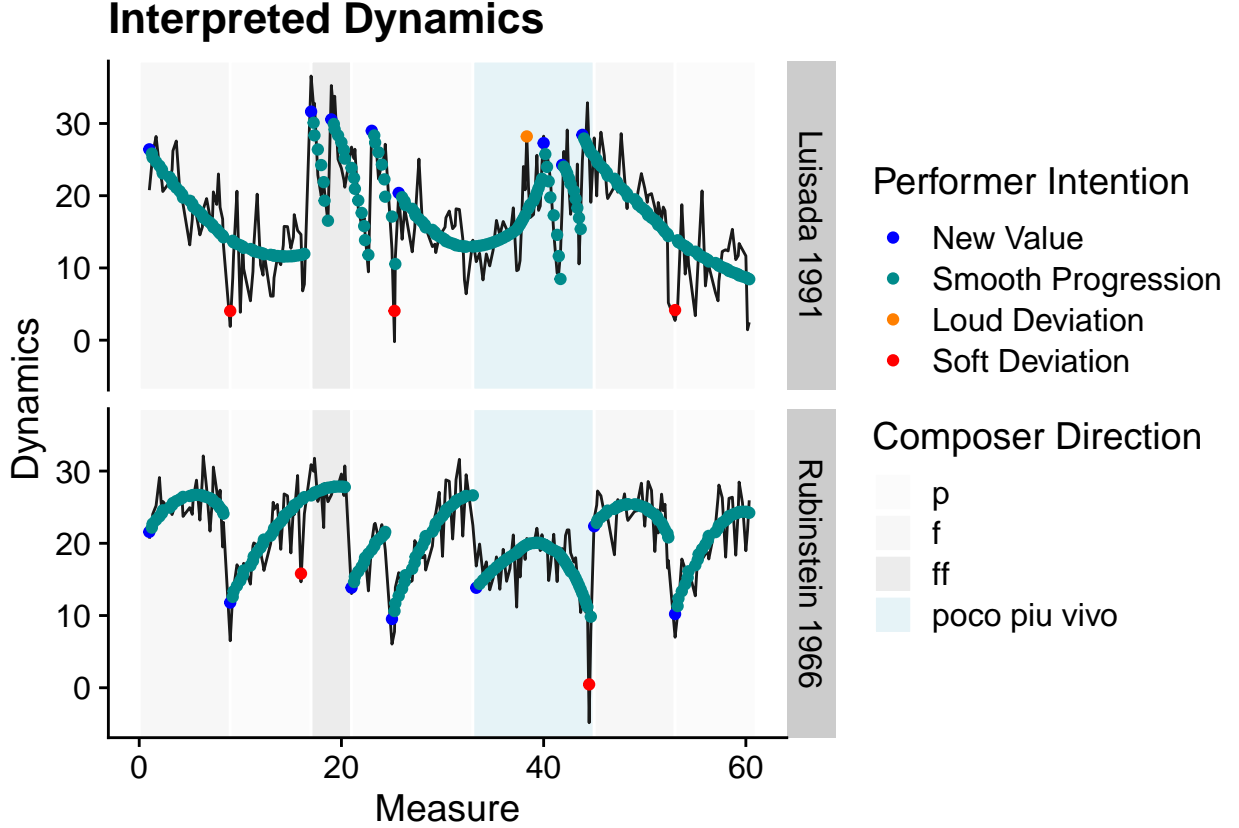
Figure 3: Interpreted Dynamics from selected performances of Chopin's Mazurka Op. 68 No. 3. The black line traces the observed dynamics while the colored dots are the smoothed interpreted dynamics.

detail in Andrew Earis's paper "An algorithm to extra expressive timing and dynamics from piano recordings" (Earis 2007).

To implement Algorithm 1 on dynamics data from performances of Chopin's Mazurka Op. 68 No. 3, we must make a few specific decisons. First, we choose prior distributions applicable specifically to Chopin's Mazurka Op. 68 No. 3, which can be found in Table 2. Secondly, regarding the Discrete Particle Filter, we selected to keep up to 500 particles with the sampling criterion of keeping the largest likelihoods. Lastly, since there are many local minima, we used 10 different starting points drawn randomly and independently from each of the prior distributions for each optimization technique (both Nelder-Mead and SANN).

Graphical representations for the performances of Jean-Marc Luisada in 1991 and Arthur Rubinstein in 1966 are shown in Figure 3. These along with plots of the other

Table 3: Parameter Estimates for Selected Performances of Chopin's Mazurka Op.68 No.3

	μ_c	μ_e	σ_ϵ^2	μ_0	μ_1	μ_2	σ_0^2	σ_1^2	σ_2^2
Block 1995	8.812	-9.777	4.946	17.036	0.121	-0.075	15.546	0.215	0.006
Grinberg 1951	10.038	-9.342	5.906	19.381	0.553	-0.094	16.676	0.396	0.006
Hatto 1993	9.871	-10.167	6.524	10.913	0.050	-0.042	9.900	0.359	0.002
Luisada 1991	10.455	-9.949	13.068	27.312	-0.165	-0.110	12.098	1.612	0.067
Richter 1976	9.541	-10.499	4.399	23.422	-0.346	-0.001	16.294	0.260	0.003
Rubinstein 1966	10.771	-10.065	8.091	12.840	0.470	-0.033	17.045	0.401	0.000
Sofronitsky 1949	9.944	-10.538	9.154	27.735	-0.234	-0.070	13.768	0.659	0.019

forty-four performances can be found in Appendix C. The black line traces the observed dynamics for each note across musical time as given by the score. The colored dots are used to differentiate the 4 possible discrete states and show the interpreted musical dynamics for each note. The background is shaded based on musical directions given by the composer in the score. Notice with these performances, the model often seems to start a new smooth progression when the composer gives direction about the dynamics. There are occasions where this is not always the case. Notice that Luisada appears to smoothly adjust the dynamics from the second *piano* section into the *poco piu vivo* section and again when transitioning from the final *forte* section to the final *piano* section. Similarly, Rubinstein also smoothly transitions but he does this between the first *forte* section into the only *fortissimo* section. These unscripted parts of the performance are what we hope to discover as they allow us to contrast and compare each performance.

Analyzing the graphical representations of each performance have the benefit of illuminating complexities within a performance but may be cumbersome when attempting to compare and contrast many performances. Instead, we analyze the parameter and transition probability values, Θ. All estimated values for the 46 performances can be found in Appendix A and Appendix B. For the reader's convenience, the parameter estimates for performances discussed in this section can be found in Table 3 and Table 4. Along with the tables of estimated parameters, we also supply density plots of the estimated parameters. These density plots show the distribution of the estimated parameters and should not be

Table 4: Probability Estimates for Selected Performances of Chopin's Mazurka Op.68 No.3

	p_{21}	p_{23}	p_{24}	p_{41}
Block 1995	0.101	0.027	0.025	0.124
Grinberg 1951	0.067	0.027	0.027	0.077
Hatto 1993	0.056	0.070	0.067	0.053
Luisada 1991	0.037	0.020	0.024	0.136
Richter 1976	0.046	0.017	0.051	0.044
Rubinstein 1966	0.034	0.017	0.023	0.214
Sofronitsky 1949	0.060	0.015	0.019	0.101

mistaken for the estimated posterior distributions for some performance.

We analyze the density plots in order to get a better understanding of the distribution of the parameters. We begin with a look at the estimated parameters, μ_0 and μ_1. The parameter μ_0 can be interpreted as the most likely starting point for each smoothed quadratic section. Joyce Hatto's 1993 performance is estimated to have the lowest average starting point with $\mu_0 = 10.913$ and Vladimir Sofronitsky's 1949 performance is estimated to have the highest average starting point with $\mu_0 = 27.74$. The average of all the μ_0's across all 46 performances is `19.148`. Figure 4 displays the density plot of μ_0 which reveals the distribution of μ_0's is fairly symmetric but with more concentration around the mean than the normal distribution. The parameter μ_1 can be interpreted as the average initial change in the dynamics of a performance. As long as μ_2 is small relative to μ_1 or if they both have the same sign, this change will persistent through each smoothed quadratic section. The parameter estimates for these performances range from a low of -0.346 in Sviatoslav Richter's 1976 performance to a high of 0.553 in Maria Grinberg's 1951 performance. From the density plot for μ_1, the distribution of the estimated parameters appears roughly normal with a mean equal to `0.103`. The mean being greater than zero indicates that, on average, performers tend to (at least initially) increase the volume through each smoothed quadratic section.

In order to get a fuller understanding of the parameters, we compare the estimated parameters in Luisada's 1991 performance with Rubinstein's 1966 performance. Luisada's
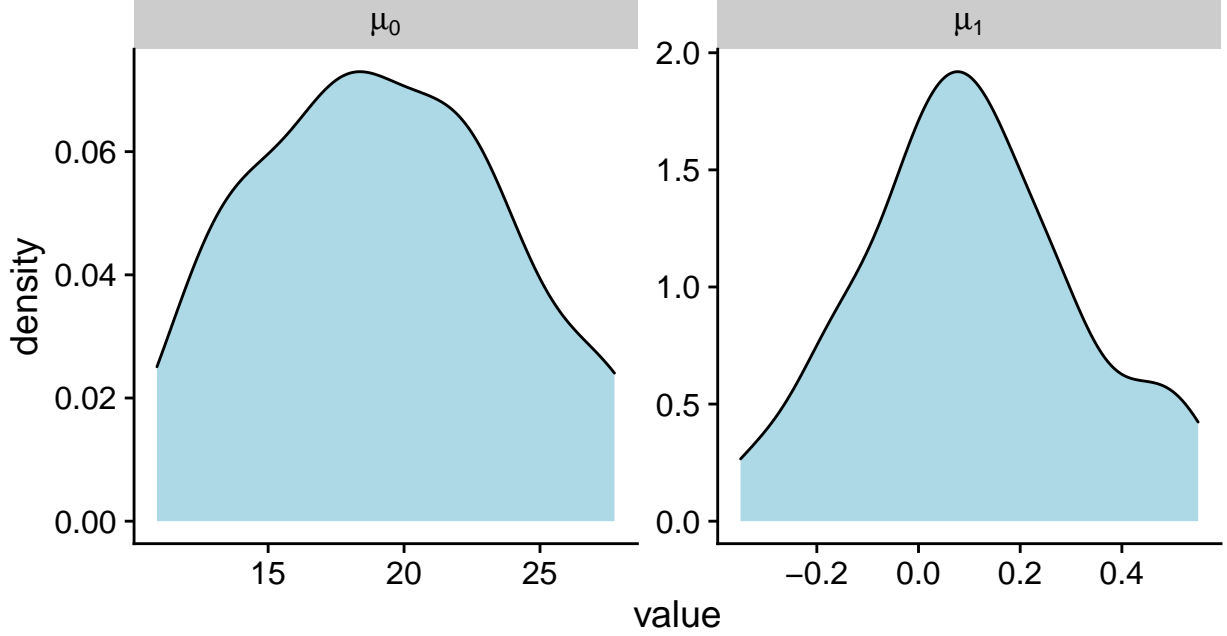
13

Figure 4: Density Plots of Estimated Parameters for 46 Performances of Chopin's Mazurka Op. 68 No. 3.

1991 performance has $\mu_0 = 27.312$, $\mu_1 = -0.165$, and $\mu_2 = -0.110$ indicating that he typically starts loudly and then softens. On the other hand, Rubinstein's performance has $\mu_0 = 12.840$, $\mu_1 = 0.470$ and $\mu_2 = -0.033$, indicating that he typically starts each smoothed section more softly and then gets louder. Taking an average across all note dynamics for the Luisada 1991 and Rubinstein 1966 performances yields `17.33` and `20.98`, respectively. While it is interesting to know that Rubinstein's 1966 was louder, on average, than Luisada's 1991 performance, we can see that without modeling the performer's intentions, interesting details are being missed. We also analyze the variance within each performance. Calculating the overall variation of note dynamics for the Luisada 1996 performance reveals a variance of `48.39`. Calculating the overall note dynamics of the Rubinstein 1996 performance reveals more homogeneity with a variance of `31.84`. Although the variation in note dynamics is higher for the Luisada 1996 performance, we observe through modelling the performer's intentions that Luisada typically has less variation in the starting dynamics of each piece-wise quadratic section than Rubinstein with $\sigma_0^2 = 12.098$ and 17.045, respectively. This is a new characteristic of the performance that would be otherwise hidden without using the music dynamics model.

14

Lastly, we compare the four estimated transition probabilities. The estimates for p_{21} and p_{41}[1] show how likely is it for the performer to start a new smoothed section. The model also reveals how likely the performer will emphasize a note by playing it more loudly through the estimated value p_{23}. For the Luisada 1991 and Rubinstein 1966 performances, all of the probabilities are relatively the same, but we do find differences in some of the other performances. For example, while the Luisada 1991 and Rubinstein 1966 performances have relatively low probabilities in starting a new smoothed section, Michel Block's 1995 performance is more likely with $p_{21} = .101$. Also, the Block 1995 performance is more likely to have emphasized notes with estimated $p_{23} = 0.027$, a probability larger than either Luisada and Rubinstein's performances.

4   Conclusion

Using a Markov-switching state space model, we were able to illuminate the intentions of a classical music performance with regards to the dynamics. In the process, a series of parameters were estimated that can be perceived as musical attributes hidden in the data. Using these attributes, we discussed a few pieces and showed various characteristics that would have been hidden by simply sticking to commonly used statistics like means or standard deviations. We then proceeded to make contrasts and comparisons of the various pieces based on these musical attributes.

The analysis described here can be taken a step further. In McDonald et al. (2019), the authors chose to cluster the performances based on the estimated parameters from their tempo model using hierarchical clustering. With the estimated parameters here and/or combining them with the estimated parameters of their tempo model, one could use their favorite clustering or classifying techniques on the performances. With additional knowledge from the data, it seems likely that the ability to cluster or classify would be improved.

Clustering and classifying are the foundational methods used in music recommendation systems that many users of music media platforms rely in order to discover or find new music. For example, the fast growing online music streaming service, Spotify, has a "Made For You" section that creates recommended playlists for the user based on their

---

[1]We focus on p_{21} more than p_{41} as being in state 4 is much less likely.

listening habits. In order to provide these recommendations, Spotify's algorithm takes into consideration information about which songs you "like", "share", and even "skip" along with information from other users that are deemed similar (Spotify 2019). While it is certainly advantageous to use data from users experience, more information can be obtained by focusing on the musical attributes of a performance itself. Pandora, another music streaming service, created the Music Genome Project where a team of musicologists listen and analyze music in order to assign 450 musical attributes (Pandora 2020). These 450 "genes" allow the streaming company to classify or cluster musical pieces in order to make recommendations to its users. It would be interesting to see if and how these additional parameters can improve the ability to classify performances and ultimately create better musical recommendation systems.

Appendices

A   Estimates of Parameters for each Performance

	μ_c	μ_e	σ^2_ϵ	μ_0	μ_1	μ_2	σ^2_0	σ^2_1	σ^2_2
Bacha 1997	10.07	-9.89	9.08	18.80	-0.01	-0.02	11.81	0.16	0.00
Barbosa 1983	9.72	-10.03	8.07	21.56	-0.07	-0.01	7.62	0.56	0.06
Biret 1990	9.80	-10.10	7.14	14.57	0.08	-0.02	11.50	0.19	0.00
Block 1995	8.81	-9.78	4.95	17.04	0.12	-0.07	15.55	0.21	0.01
Brailowsky 1960	10.33	-11.21	9.98	25.96	0.22	-0.03	11.85	0.10	0.00
Chiu 1999	10.29	-9.90	6.44	13.10	0.13	0.00	13.43	0.45	0.01
Clidat 1994	10.09	-9.46	6.68	21.43	0.05	0.01	12.21	0.65	0.01
Cohen 1997	10.71	-8.85	7.19	22.24	0.08	-0.05	17.08	0.28	0.01
Cortot 1951	10.04	-10.78	10.23	17.87	-0.16	0.01	11.73	0.22	0.02
Csalog 1996	10.24	-10.06	10.54	23.20	-0.07	-0.04	7.20	0.55	0.53
Czerny-Stefanska 1990	10.85	-8.98	5.52	16.52	0.27	-0.05	13.42	1.19	0.00
Ezaki 2006	10.00	-9.04	5.95	21.22	-0.15	0.01	11.26	0.13	0.00
Ferenczy 1958	10.26	-9.69	8.24	20.67	0.22	-0.02	20.81	0.12	0.00
Fliere 1977	11.48	-10.58	12.40	17.68	0.28	-0.05	16.40	0.50	0.01
Fou 1978	9.84	-10.10	7.70	16.98	-0.18	0.07	9.92	2.14	0.08
Francois 1956	9.78	-9.89	10.95	19.55	-0.05	0.02	14.47	0.33	0.01
Grinberg 1951	10.04	-9.34	5.91	19.38	0.55	-0.09	16.68	0.40	0.01
Hatto 1993	9.87	-10.17	6.52	10.91	0.05	-0.04	9.90	0.36	0.00
Hatto 2006	9.73	-10.80	5.49	11.43	0.16	-0.09	11.91	0.74	0.01
Indjic 1988	9.47	-10.16	7.99	16.69	0.00	-0.03	14.92	0.53	0.00
Jonas 1947	9.89	-10.39	8.45	13.21	0.20	-0.04	12.59	1.69	0.03
Kapell 1951	10.62	-10.49	6.85	17.87	0.52	-0.05	12.41	0.06	0.00
Kiepura 1999	10.51	-9.75	9.23	14.02	0.15	-0.04	8.41	0.96	0.00

	μ_c	μ_e	σ^2_ϵ	μ_0	μ_1	μ_2	σ^2_0	σ^2_1	σ^2_2
Kushner 1989	9.92	-9.22	7.28	13.00	0.41	-0.03	12.24	0.10	0.00
Luisada 1991	10.45	-9.95	13.07	27.31	-0.17	-0.11	12.10	1.61	0.07
Lushtak 2004	9.21	-8.70	6.23	14.01	0.13	0.10	14.54	0.17	0.02
Meguri 1997	10.32	-11.08	8.20	19.91	0.10	-0.03	13.37	1.31	0.00
Milkina 1970	9.63	-9.71	8.09	14.11	0.29	-0.02	11.50	0.50	0.00
Mohovich 1999	11.32	-9.66	7.47	22.00	-0.13	0.01	13.88	0.55	0.00
Ohlsson 1999	9.54	-9.70	7.43	22.44	-0.30	-0.01	11.20	0.96	0.01
Olejniczac 1990	11.03	-9.00	5.74	17.27	0.14	0.02	14.22	0.32	0.00
Rangell 2001	9.92	-11.68	7.60	15.61	0.29	-0.08	9.49	0.49	0.02
Richter 1976	9.54	-10.50	4.40	23.42	-0.35	0.00	16.29	0.26	0.00
Rubinstein 1939	9.71	-8.83	6.42	22.56	-0.02	-0.10	10.31	1.46	0.01
Rubinstein 1952	9.79	-10.36	9.50	21.26	-0.02	-0.01	9.83	1.32	0.01
Rubinstein 1961	9.74	-9.89	6.95	13.66	0.50	-0.04	13.10	0.31	0.00
Rubinstein 1966	10.77	-10.06	8.09	12.84	0.47	-0.03	17.05	0.40	0.00
Shebanova 2002	10.20	-10.40	10.24	18.33	0.03	0.00	5.54	0.08	0.00
Smidowicz 1948a	9.97	-9.99	7.86	18.18	0.32	-0.06	7.97	0.27	0.00
Smidowicz 1948b	10.07	-9.47	7.61	27.62	0.22	-0.03	13.90	0.06	0.00
Smith 1975	10.51	-9.57	9.21	27.66	0.02	-0.10	10.95	1.12	0.02
Sofronitsky 1949	9.94	-10.54	9.15	27.74	-0.23	-0.07	13.77	0.66	0.02
Sztompka 1959	9.73	-9.10	6.55	23.33	0.01	-0.01	10.74	0.63	0.01
Tomsic 1995	9.44	-9.22	4.62	18.17	0.41	-0.11	14.28	1.19	0.03
Uninsky 1971	9.93	-10.87	7.42	23.89	0.12	-0.06	10.56	1.28	0.01
Wasowski 1980	10.31	-9.73	7.26	24.58	0.07	-0.01	9.85	0.12	0.00
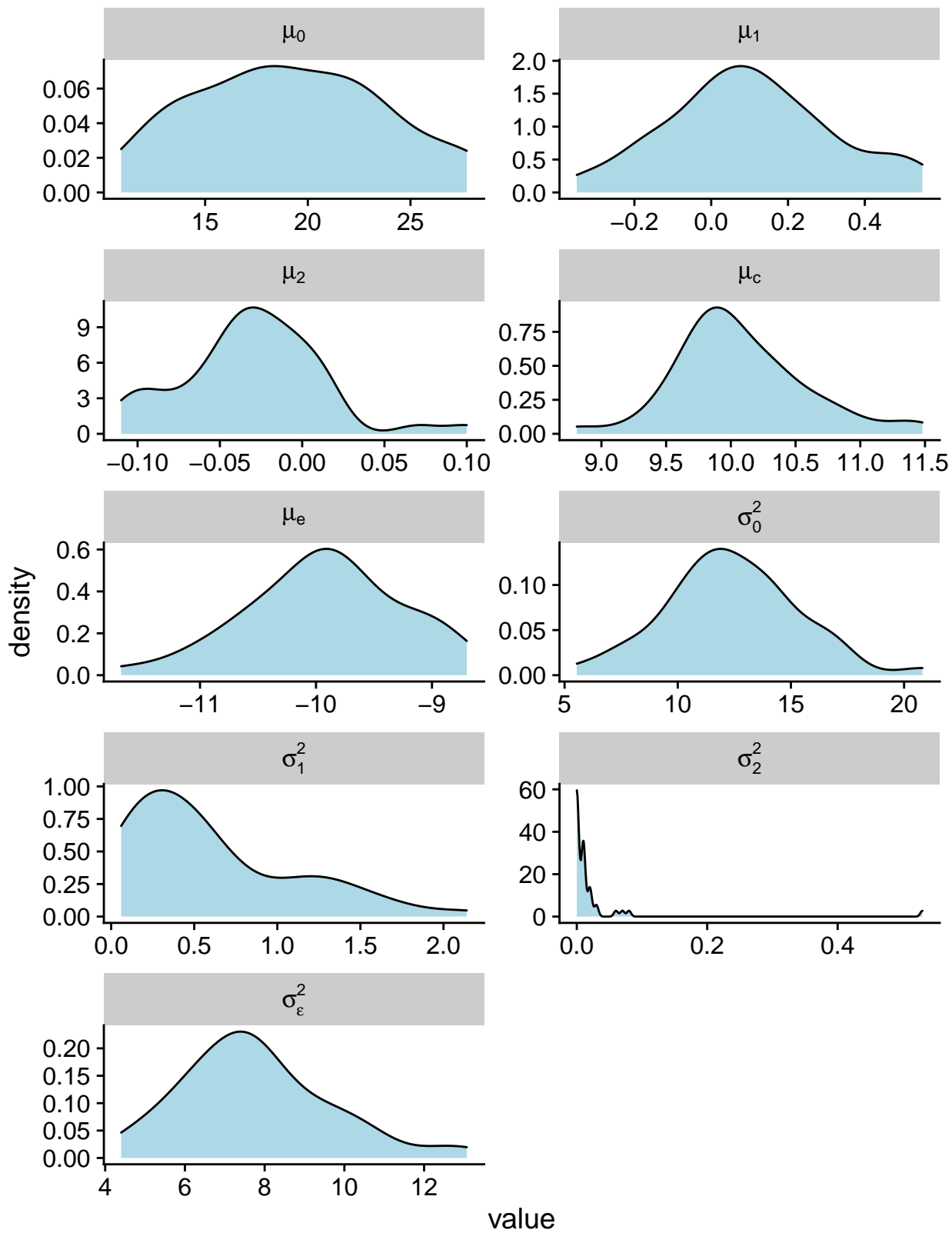
B Estimates of Probabilities for each Performance

	p_{21}	p_{23}	p_{24}	p_{41}
Bacha 1997	0.03	0.02	0.03	0.08
Barbosa 1983	0.02	0.02	0.02	0.12
Biret 1990	0.03	0.04	0.03	0.12
Block 1995	0.10	0.03	0.03	0.12
Brailowsky 1960	0.05	0.03	0.09	0.04
Chiu 1999	0.04	0.02	0.03	0.07
Clidat 1994	0.05	0.02	0.04	0.06
Cohen 1997	0.05	0.02	0.05	0.08
Cortot 1951	0.03	0.01	0.06	0.29
Csalog 1996	0.02	0.02	0.01	0.14
Czerny-Stefanska 1990	0.09	0.02	0.02	0.15
Ezaki 2006	0.02	0.02	0.06	0.08
Ferenczy 1958	0.03	0.04	0.05	0.08
Fliere 1977	0.03	0.01	0.02	0.21
Fou 1978	0.05	0.02	0.03	0.09
Francois 1956	0.03	0.03	0.02	0.08
Grinberg 1951	0.07	0.03	0.03	0.08
Hatto 1993	0.06	0.07	0.07	0.05
Hatto 2006	0.05	0.04	0.07	0.05
Indjic 1988	0.03	0.07	0.06	0.04
Jonas 1947	0.06	0.04	0.02	0.10
Kapell 1951	0.06	0.02	0.03	0.05
Kiepura 1999	0.03	0.02	0.02	0.14

	p_{21}	p_{23}	p_{24}	p_{41}
Kushner 1989	0.04	0.01	0.04	0.19
Luisada 1991	0.04	0.02	0.02	0.14
Lushtak 2004	0.08	0.06	0.04	0.05
Meguri 1997	0.03	0.01	0.02	0.06
Milkina 1970	0.05	0.03	0.02	0.28
Mohovich 1999	0.05	0.02	0.02	0.11
Ohlsson 1999	0.06	0.03	0.02	0.09
Olejniczac 1990	0.05	0.02	0.07	0.07
Rangell 2001	0.04	0.01	0.06	0.05
Richter 1976	0.05	0.02	0.05	0.04
Rubinstein 1939	0.09	0.02	0.12	0.04
Rubinstein 1952	0.05	0.02	0.02	0.13
Rubinstein 1961	0.04	0.02	0.03	0.05
Rubinstein 1966	0.03	0.02	0.02	0.21
Shebanova 2002	0.05	0.03	0.07	0.03
Smidowicz 1948a	0.04	0.02	0.05	0.05
Smidowicz 1948b	0.05	0.02	0.07	0.03
Smith 1975	0.06	0.02	0.06	0.12
Sofronitsky 1949	0.06	0.01	0.02	0.10
Sztompka 1959	0.05	0.02	0.07	0.03
Tomsic 1995	0.09	0.02	0.02	0.11
Uninsky 1971	0.05	0.02	0.07	0.03
Wasowski 1980	0.03	0.02	0.02	0.14
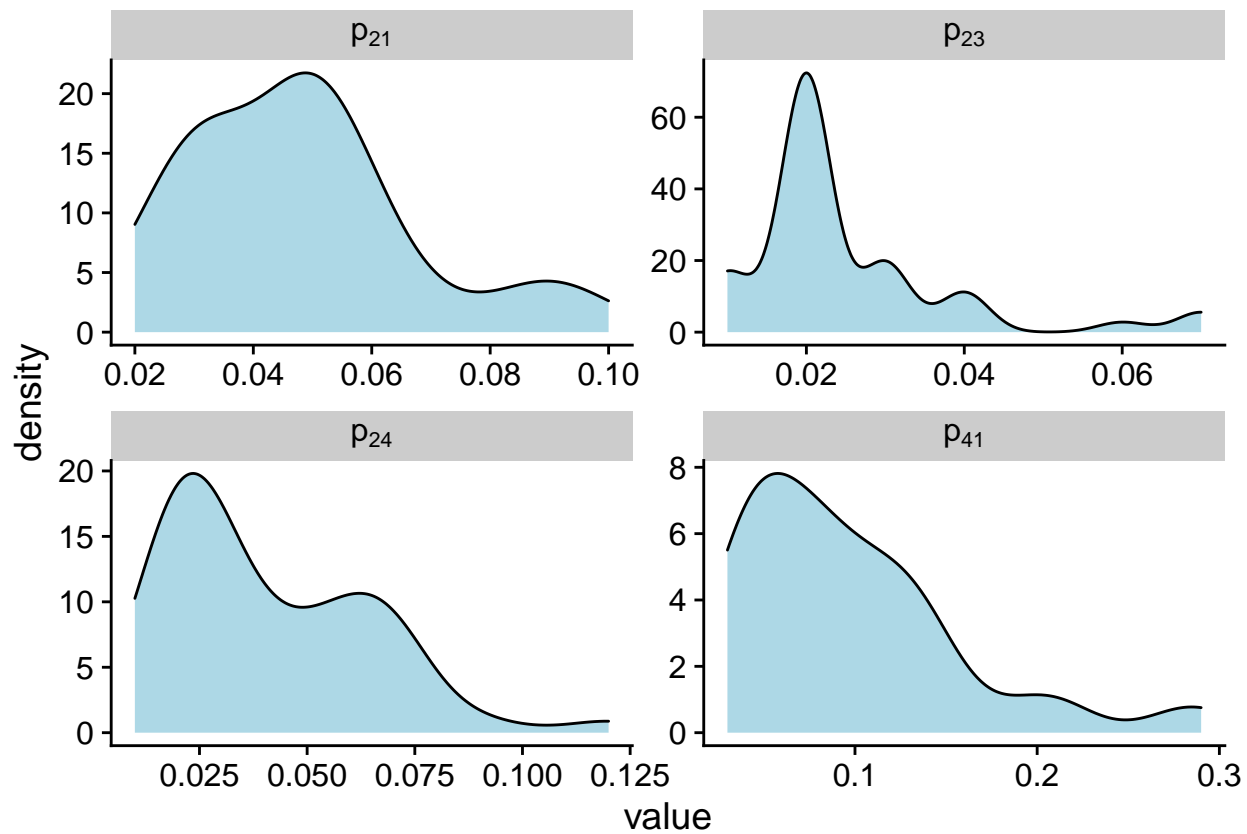
C Plots for each Performance

Bacha 1997

Barbosa 1983

Biret 1990

Block 1995

Clidat 1994

Dynamics vs Measure

Performer Intention
- New Value
- Smooth Progression
- Loud Deviation
- Soft Deviation

Composer Direction
- p
- f
- ff
- poco piu vivo

Cohen 1997

Dynamics vs Measure

Composer Direction
- p
- f
- ff
- poco piu vivo

Performer Intention
- New Value
- Smooth Progression
- Soft Deviation
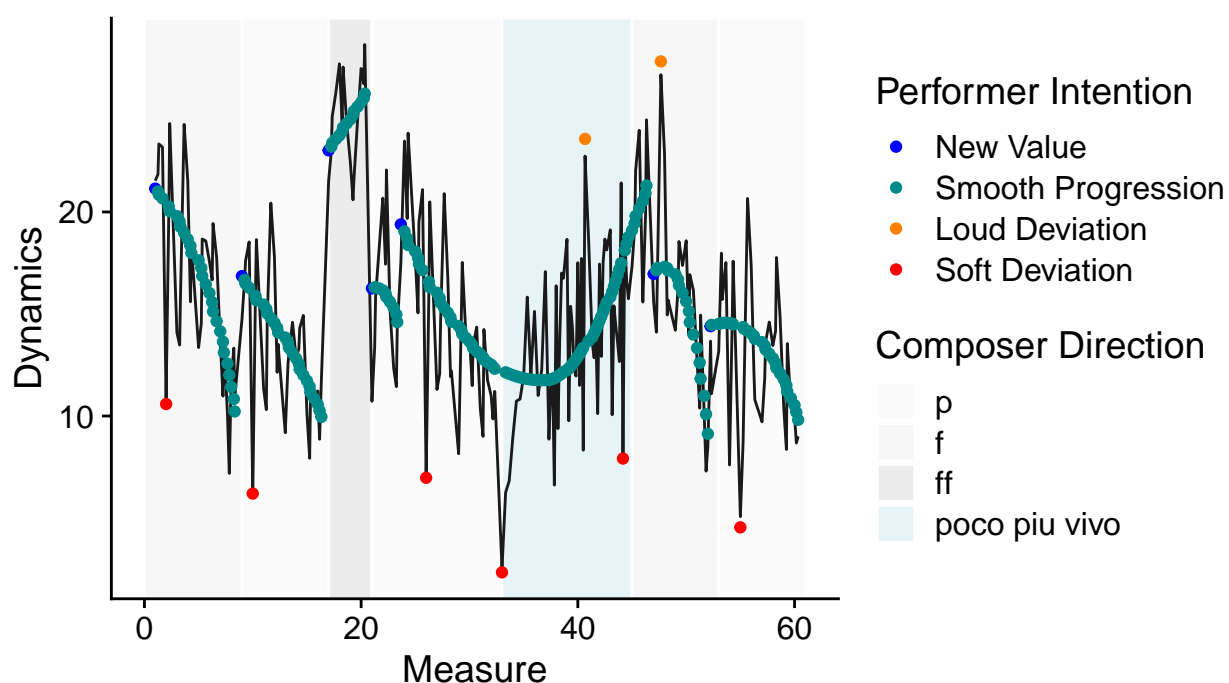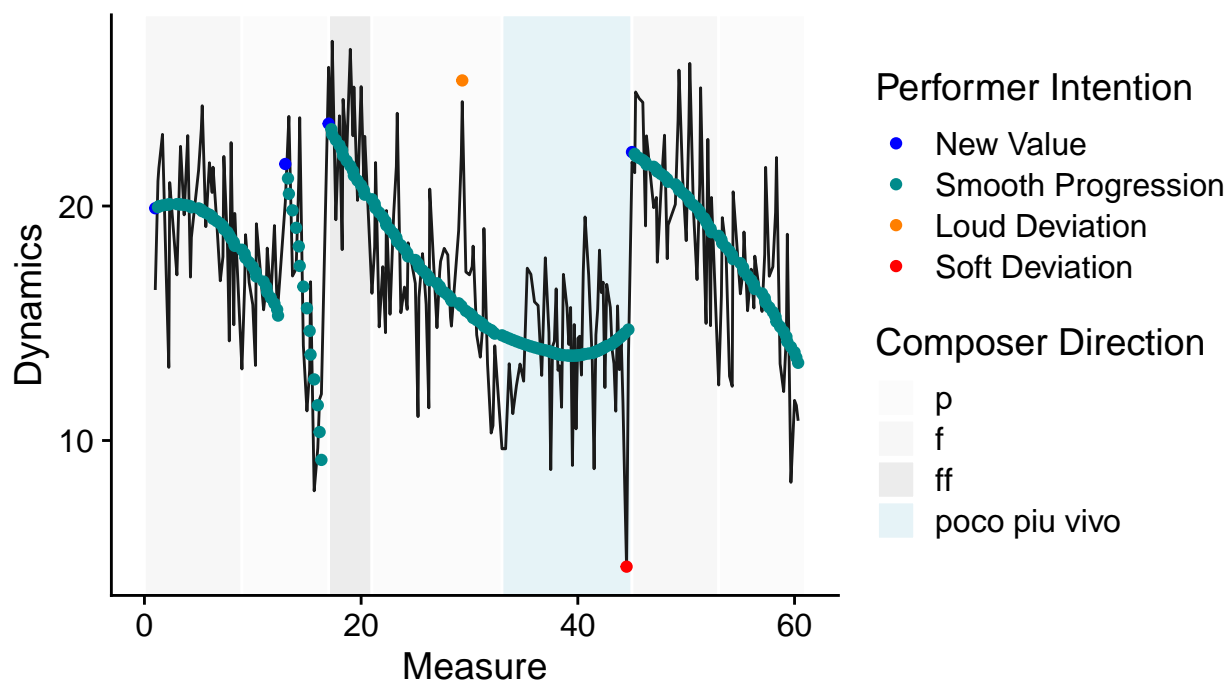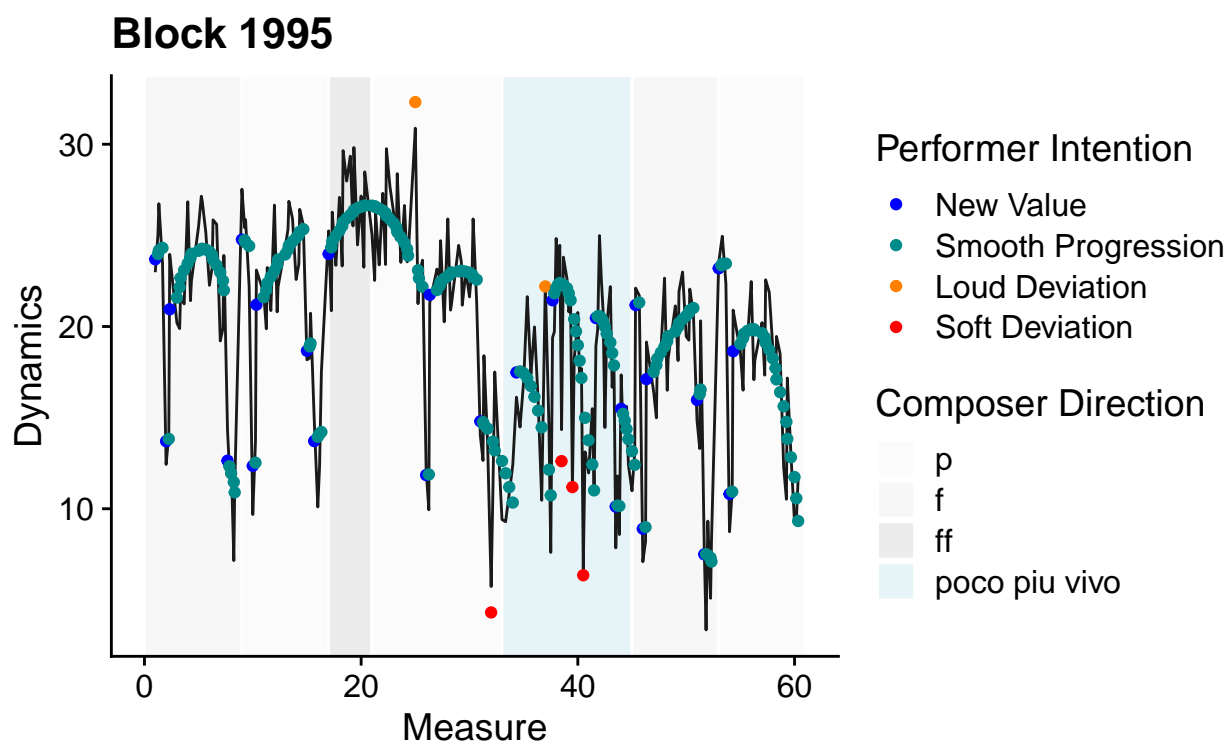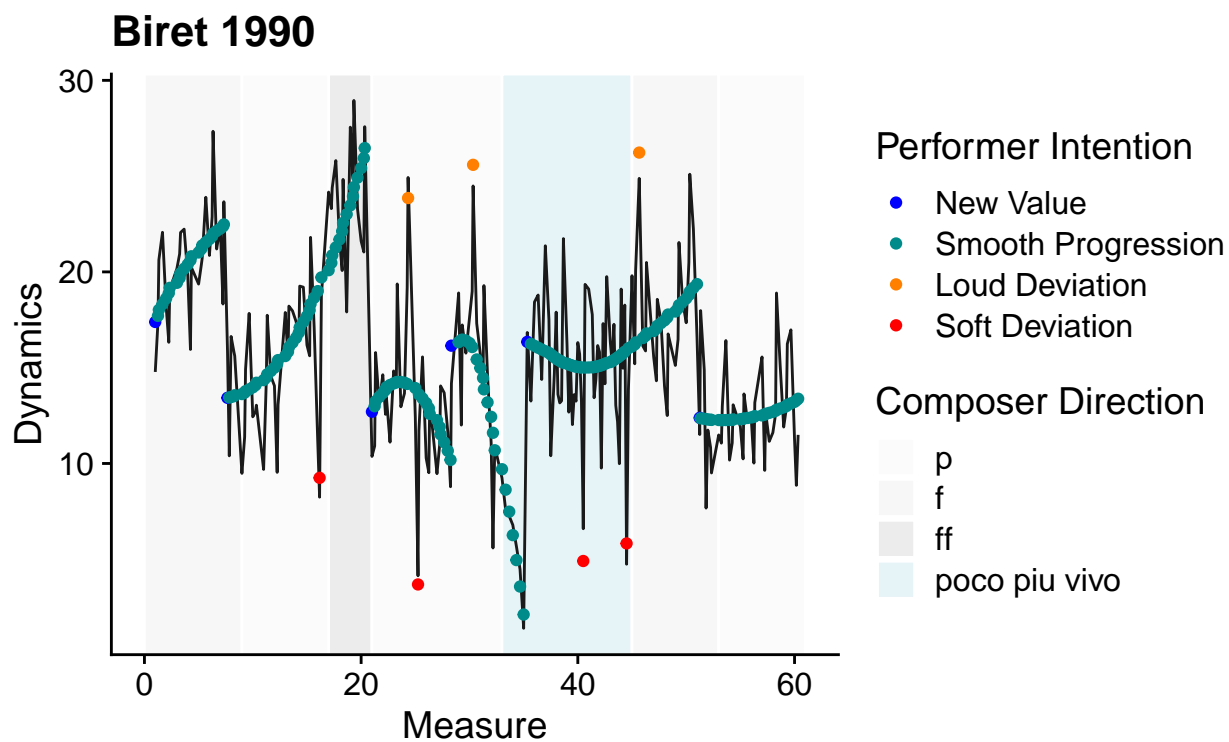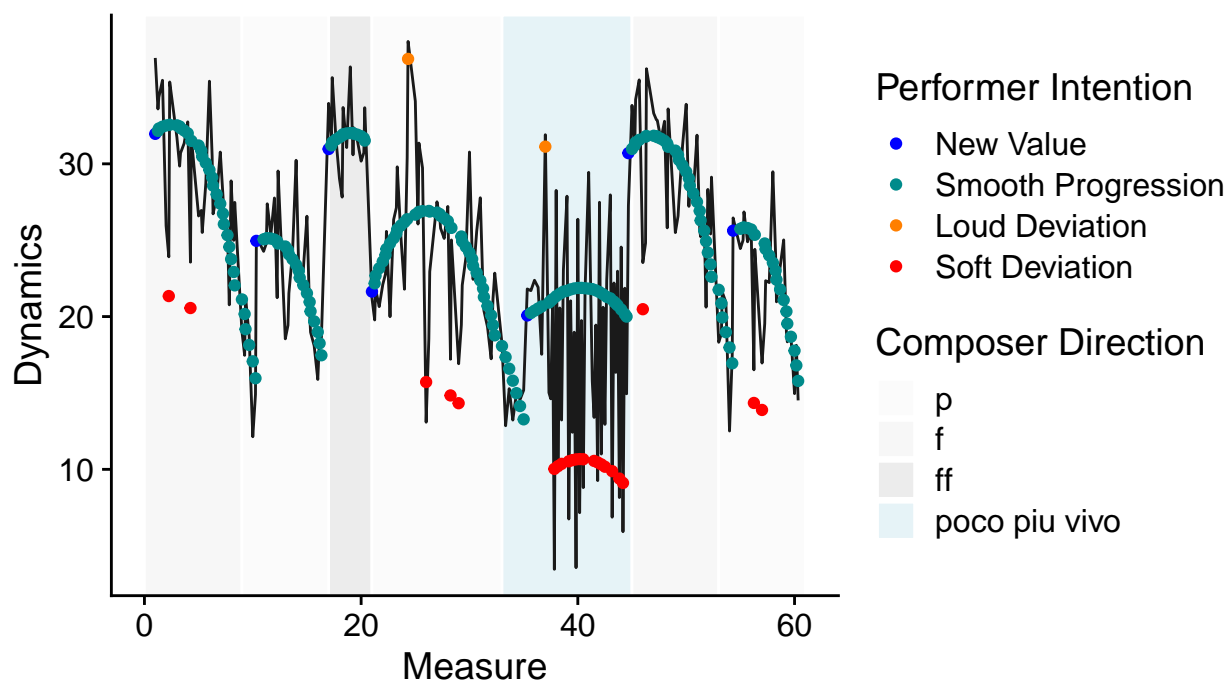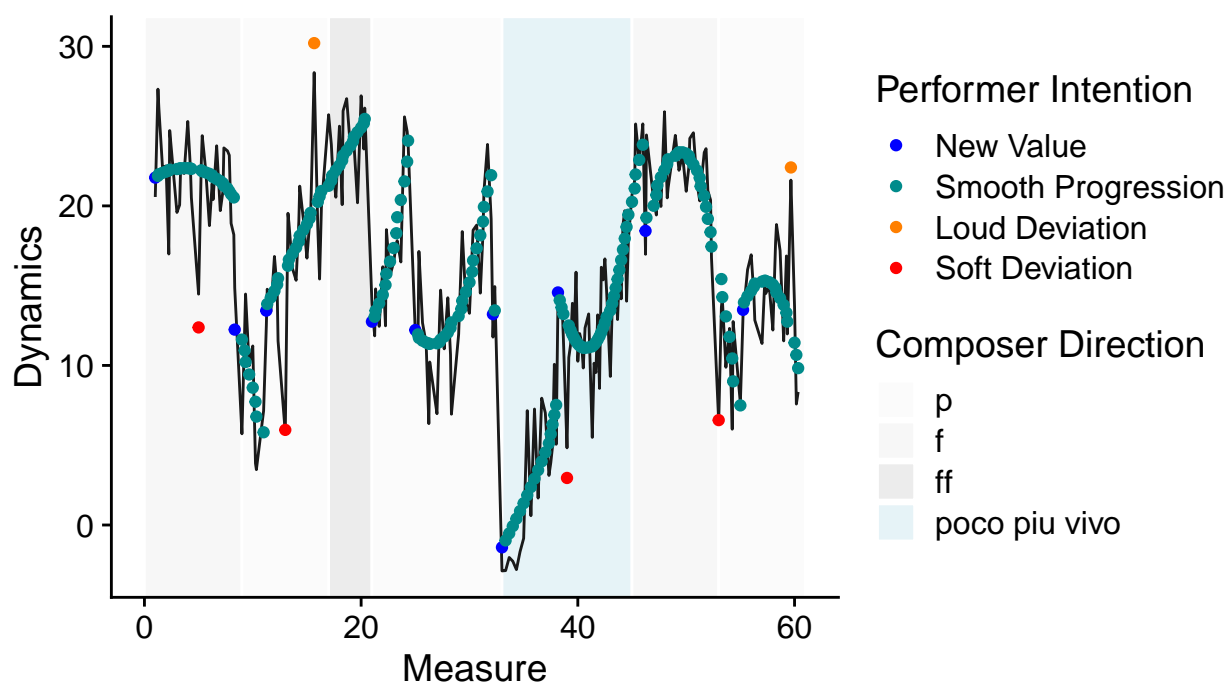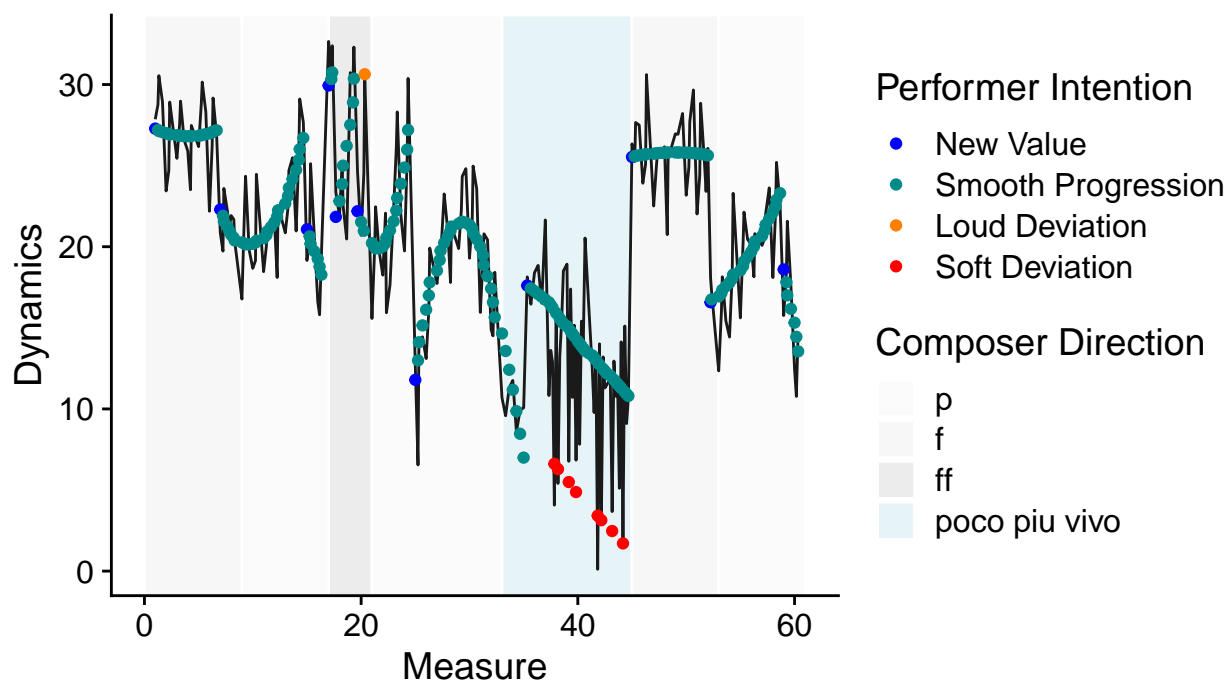
Cortot 1951

Csalog 1996

Czerny–Stefanska 1990

Ezaki 2006

Ferenczy 1958

Performer Intention
- New Value
- Smooth Progression
- Loud Deviation
- Soft Deviation

Composer Direction
- p
- f
- ff
- poco piu vivo

Fliere 1977

Composer Direction
- p
- f
- ff
- poco piu vivo

Performer Intention
- New Value
- Smooth Progression
- Soft Deviation
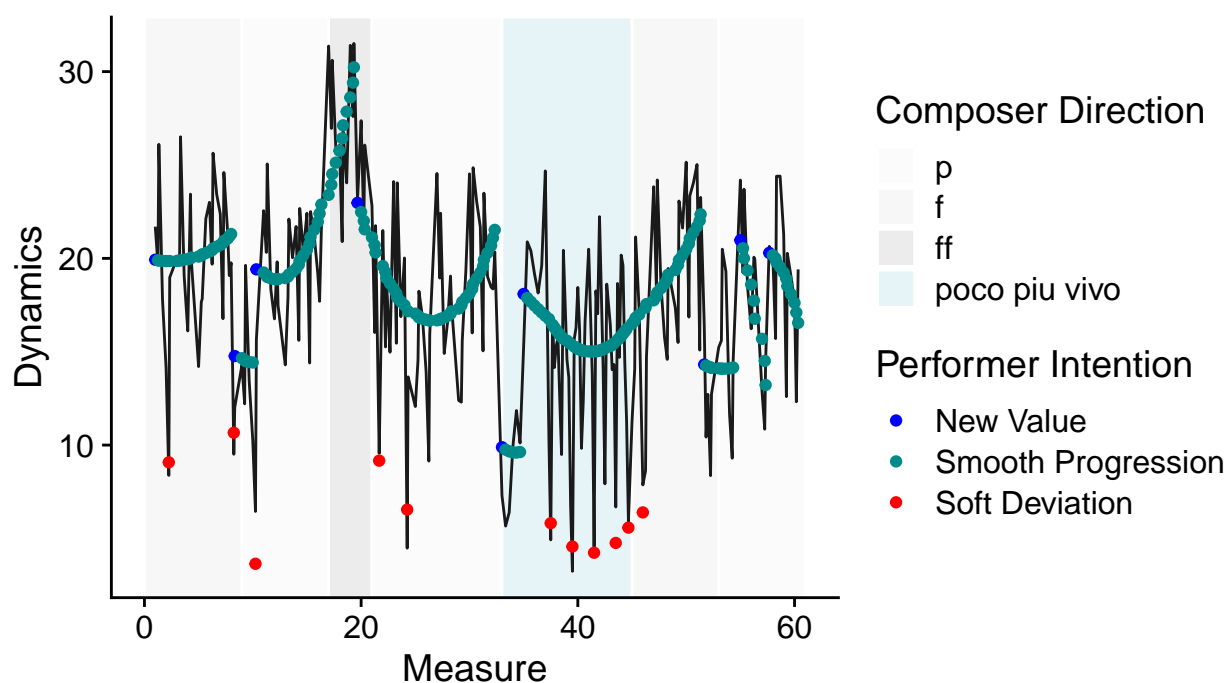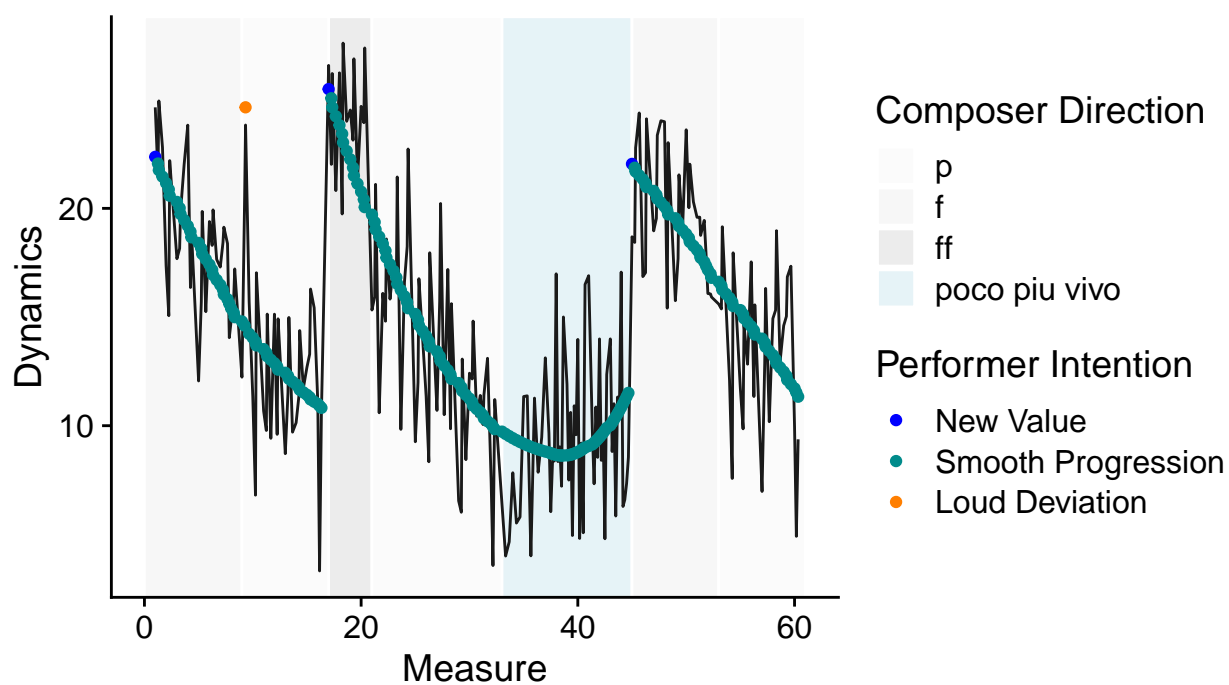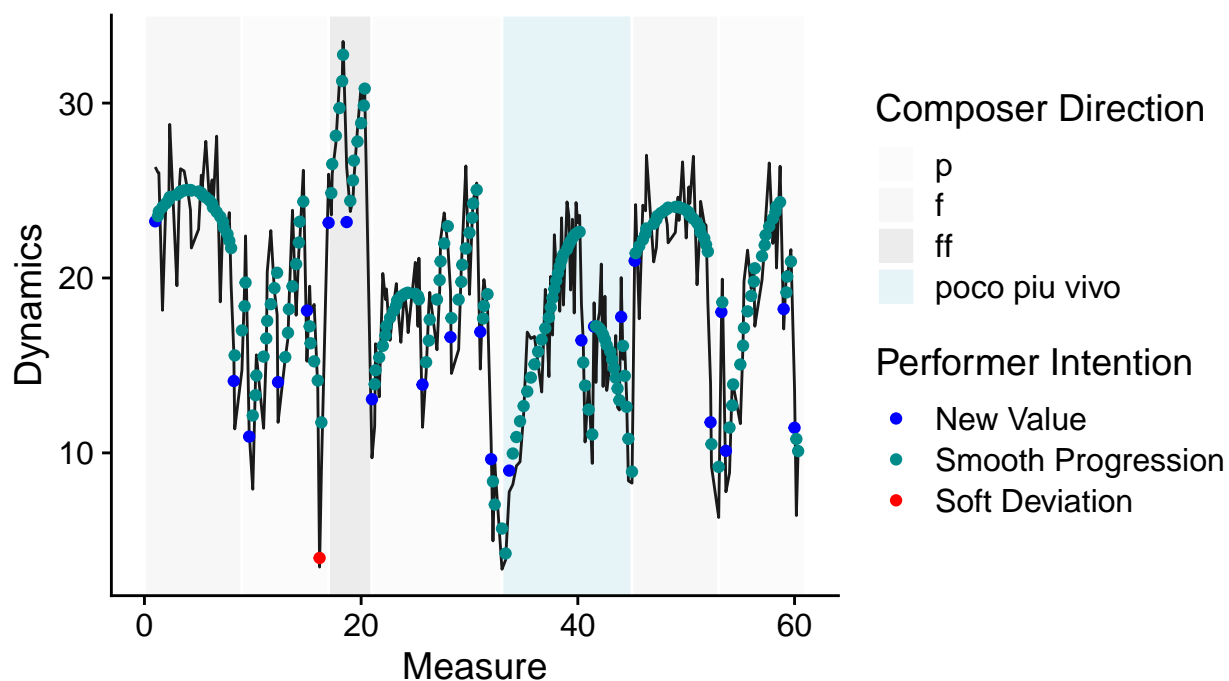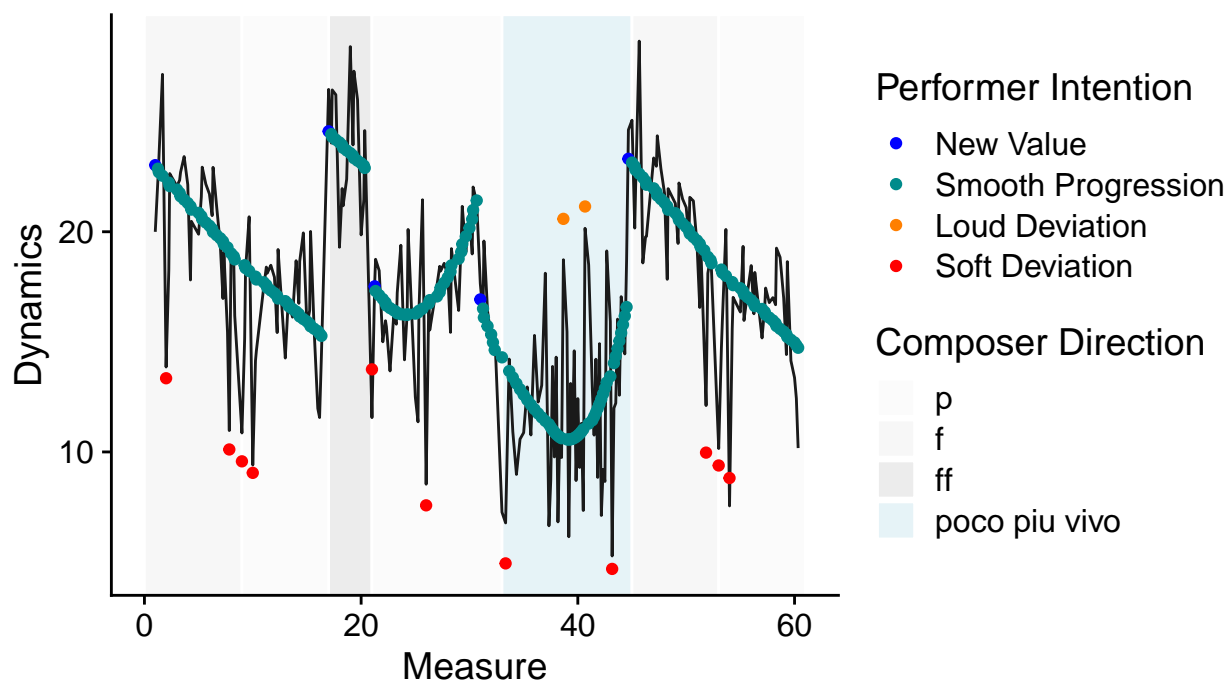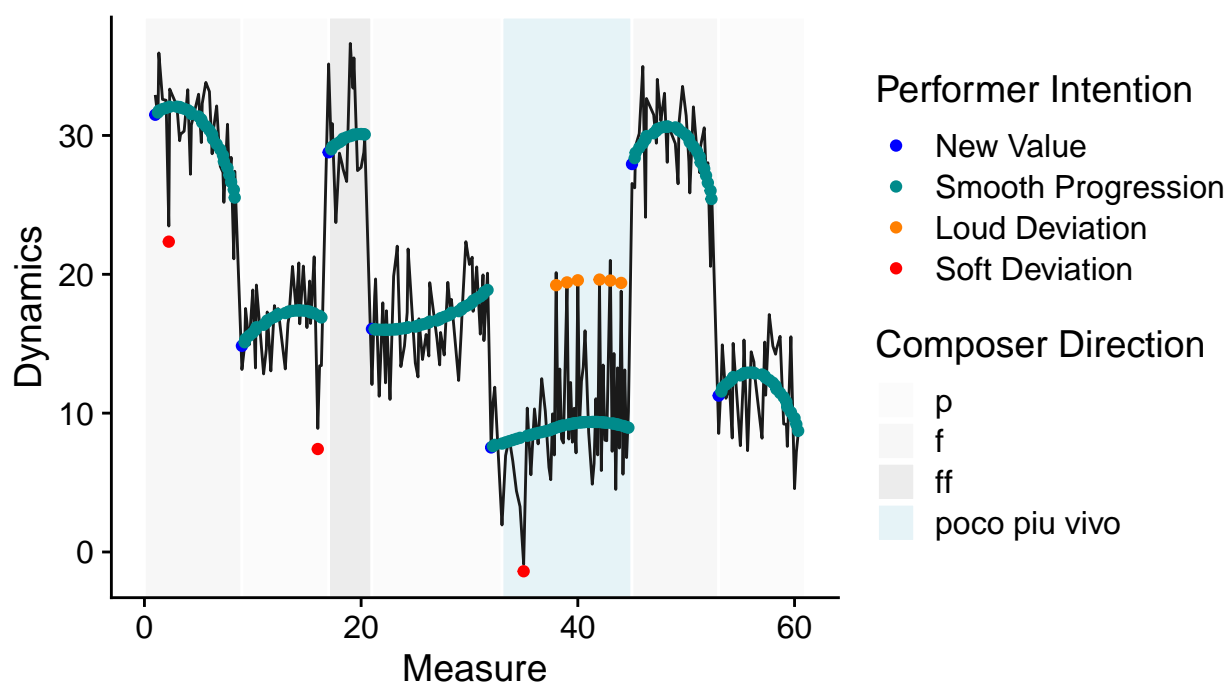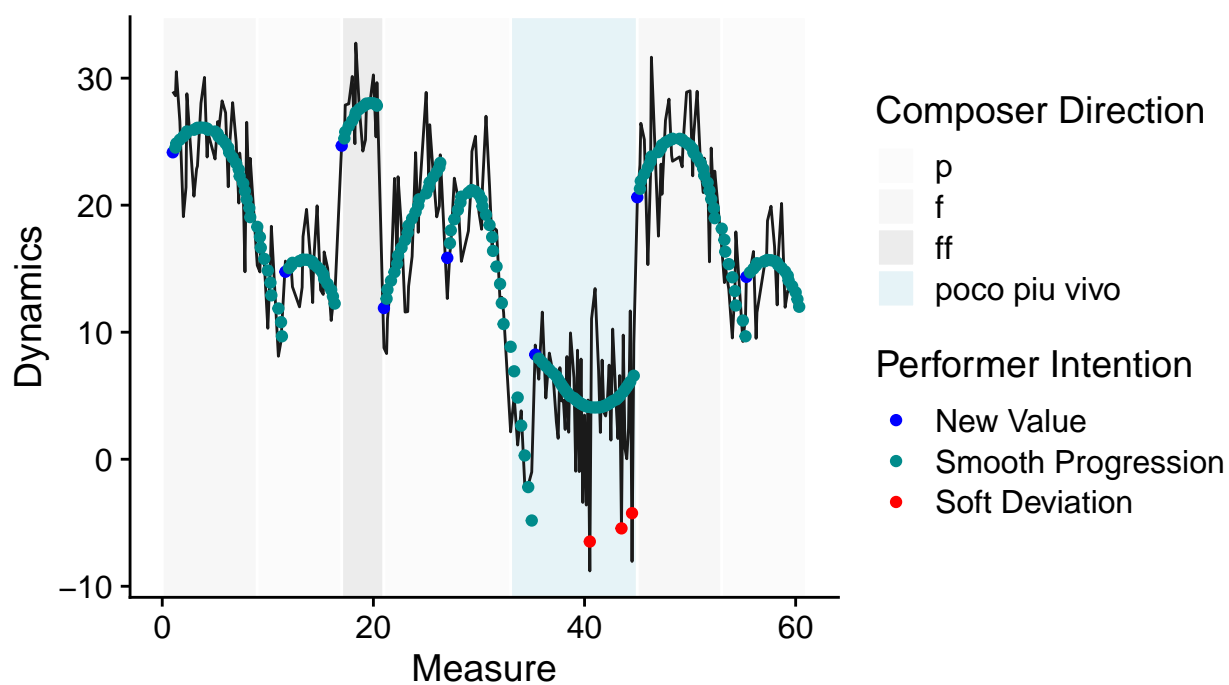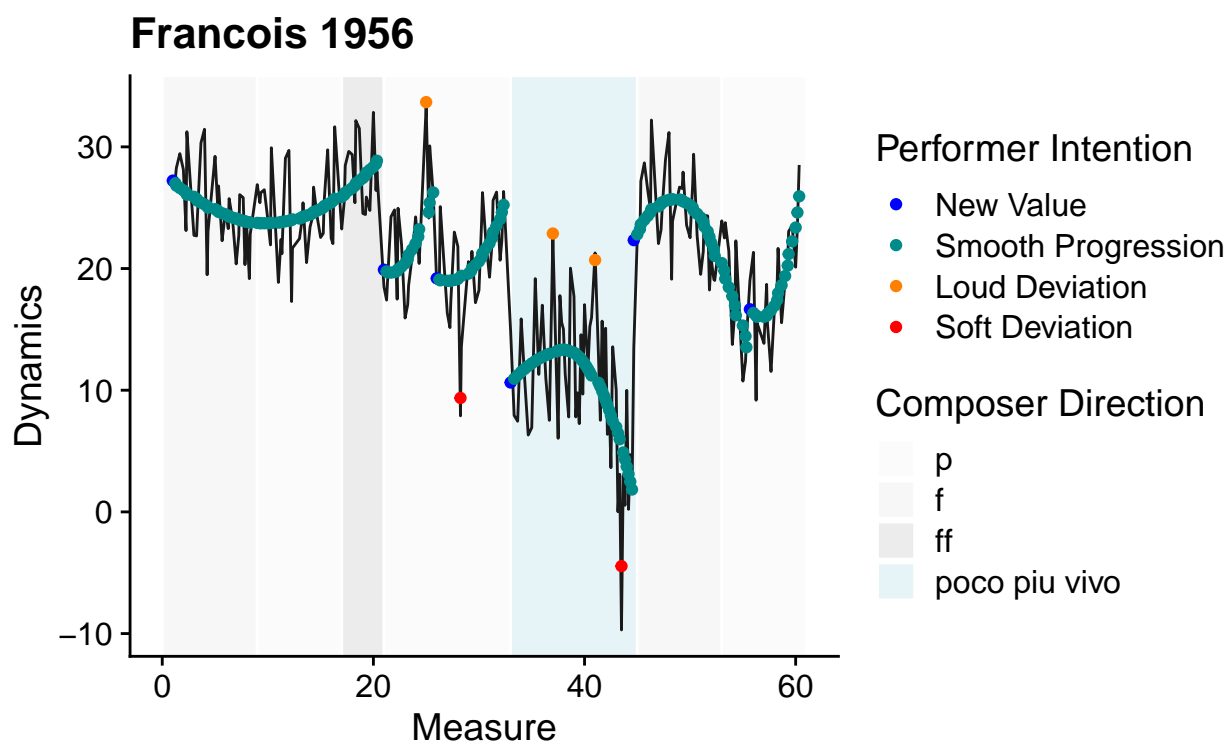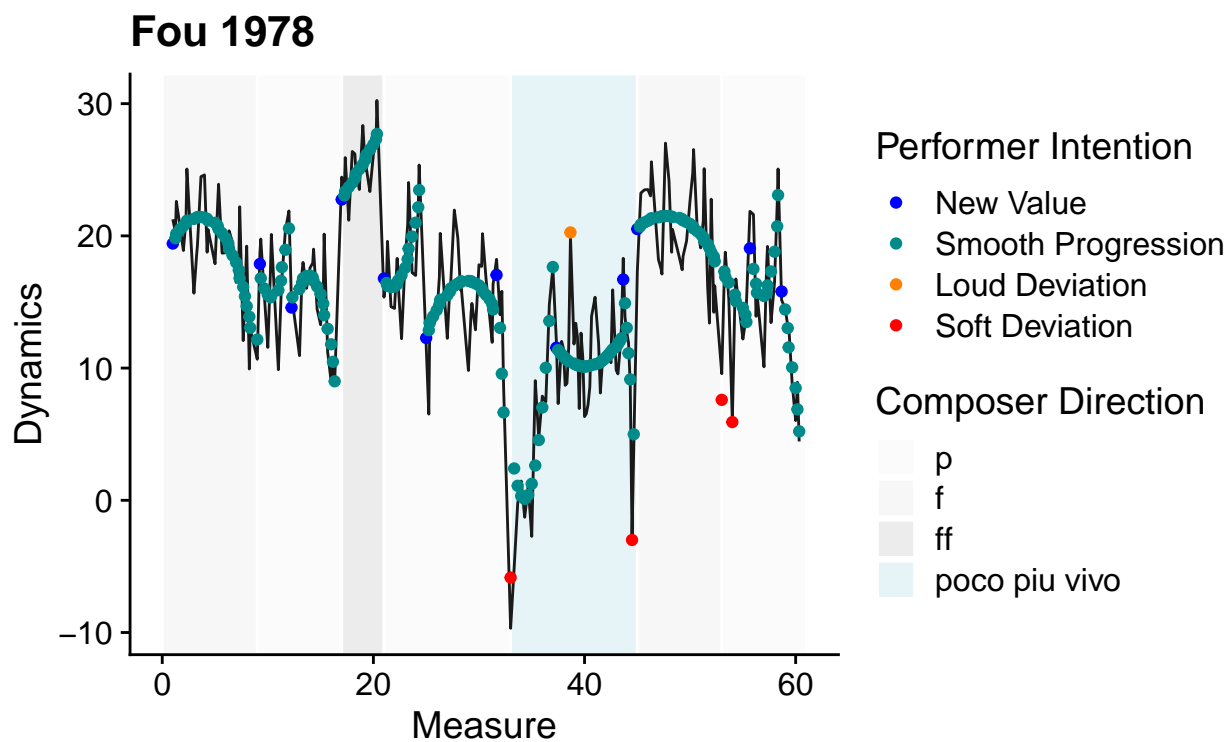
Grinberg 1951

Hatto 1993

Hatto 2006

Indjic 1988

Jonas 1947

Performer Intention
- New Value
- Smooth Progression
- Loud Deviation
- Soft Deviation

Composer Direction
- p
- f
- ff
- poco piu vivo

Kapell 1951

Performer Intention
- New Value
- Smooth Progression
- Loud Deviation
- Soft Deviation

Composer Direction
- p
- f
- ff
- poco piu vivo

Luisada 1991

Performer Intention

- New Value
- Smooth Progression
- Loud Deviation
- Soft Deviation

Composer Direction

- p
- f
- ff
- poco piu vivo

Lushtak 2004

Performer Intention

- New Value
- Smooth Progression
- Loud Deviation
- Soft Deviation

Composer Direction

- p
- f
- ff
- poco piu vivo

Meguri 1997

Performer Intention
- New Value
- Smooth Progression
- Loud Deviation
- Soft Deviation

Composer Direction
- p
- f
- ff
- poco piu vivo

Milkina 1970

Performer Intention
- New Value
- Smooth Progression
- Loud Deviation
- Soft Deviation

Composer Direction
- p
- f
- ff
- poco piu vivo

Mohovich 1999

Performer Intention
- New Value
- Smooth Progression
- Loud Deviation
- Soft Deviation

Composer Direction
- p
- f
- ff
- poco piu vivo

Ohlsson 1999

Composer Direction
- p
- f
- ff
- poco piu vivo

Performer Intention
- New Value
- Smooth Progression
- Loud Deviation

Olejniczac 1990

Composer Direction
- p
- f
- ff
- poco piu vivo

Performer Intention
- New Value
- Smooth Progression
- Soft Deviation

Rangell 2001

Composer Direction
- p
- f
- ff
- poco piu vivo

Performer Intention
- New Value
- Smooth Progression
- Soft Deviation

Richter 1976

Composer Direction
- p
- f
- ff
- poco piu vivo

Performer Intention
- New Value
- Smooth Progression
- Soft Deviation

Rubinstein 1939

Composer Direction
- p
- f
- ff
- poco piu vivo

Performer Intention
- New Value
- Smooth Progression
- Soft Deviation

Smidowicz 1948a

Smidowicz 1948b

Sztompka 1959

Performer Intention
- New Value
- Smooth Progression
- Loud Deviation
- Soft Deviation

Composer Direction
- p
- f
- ff
- poco piu vivo

Tomsic 1995

Performer Intention
- New Value
- Smooth Progression
- Loud Deviation
- Soft Deviation

Composer Direction
- p
- f
- ff
- poco piu vivo
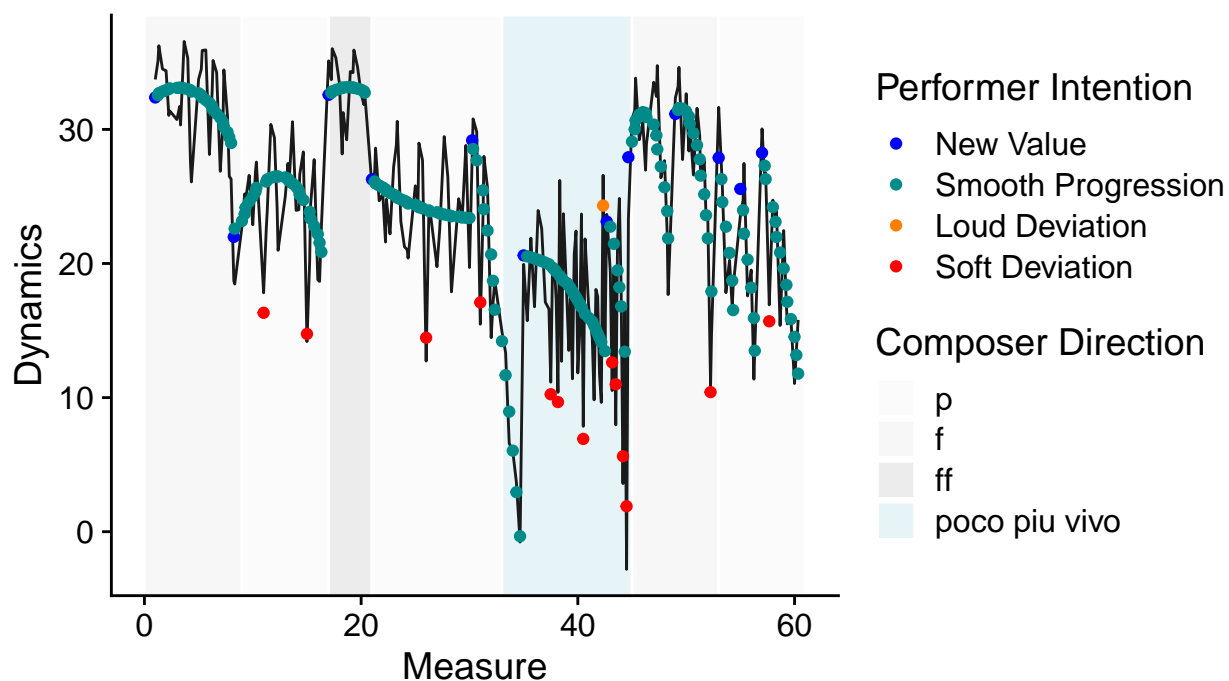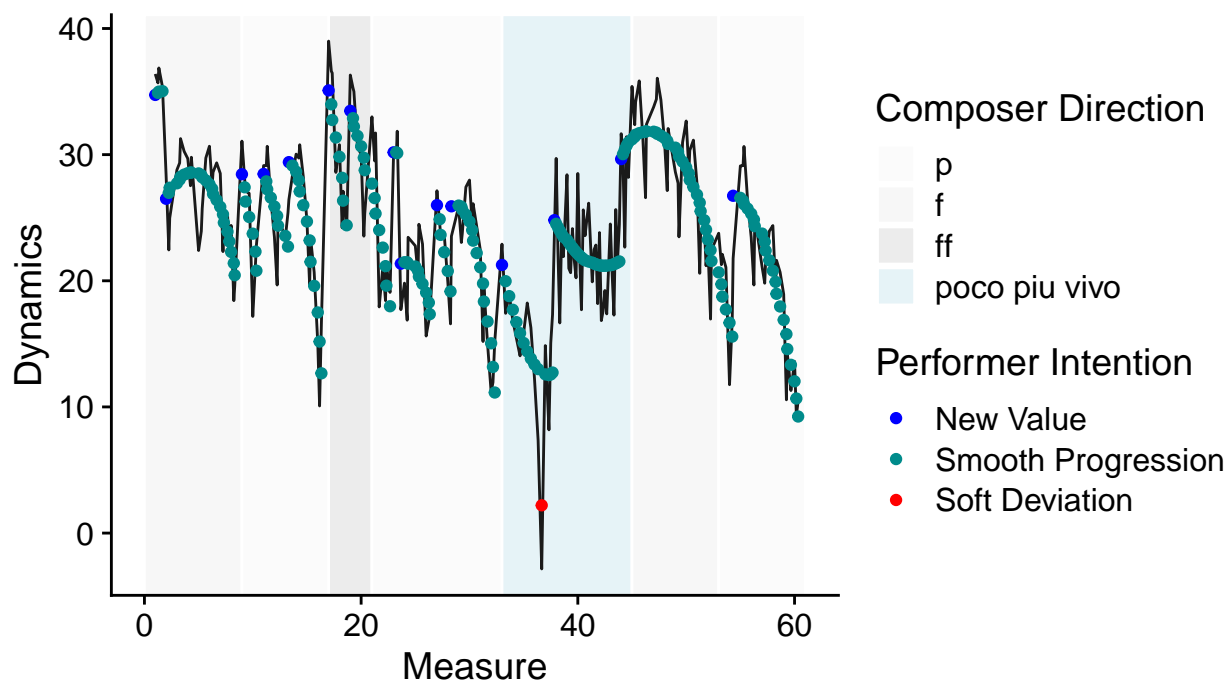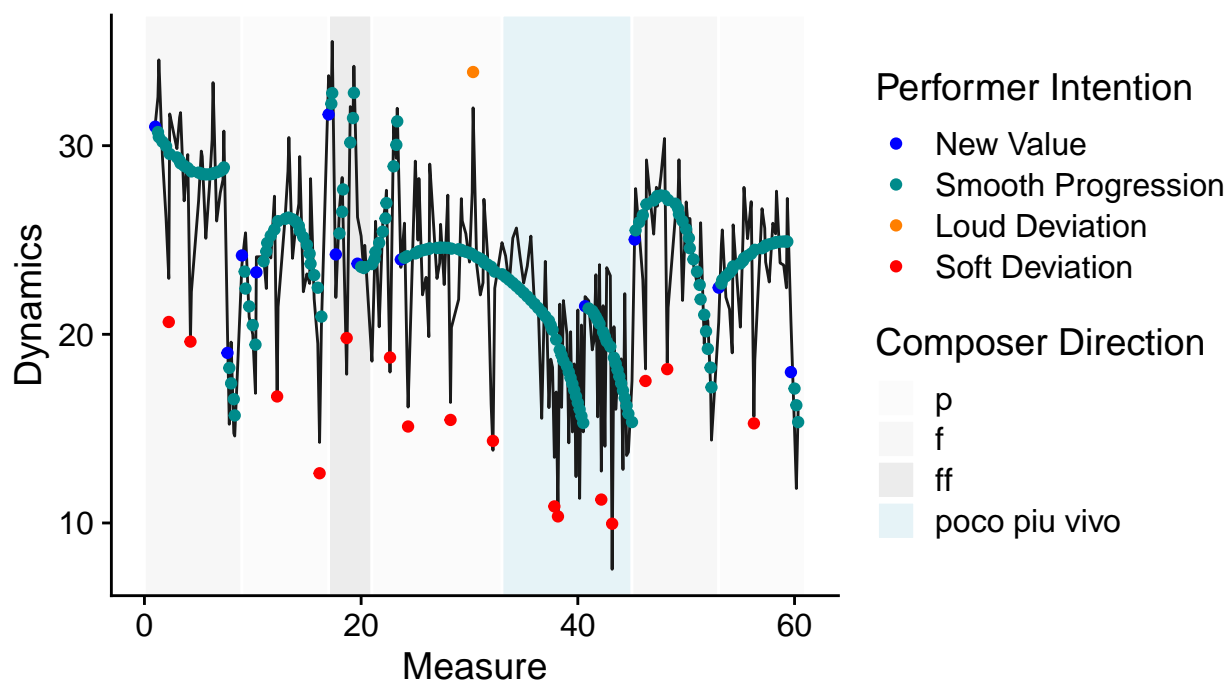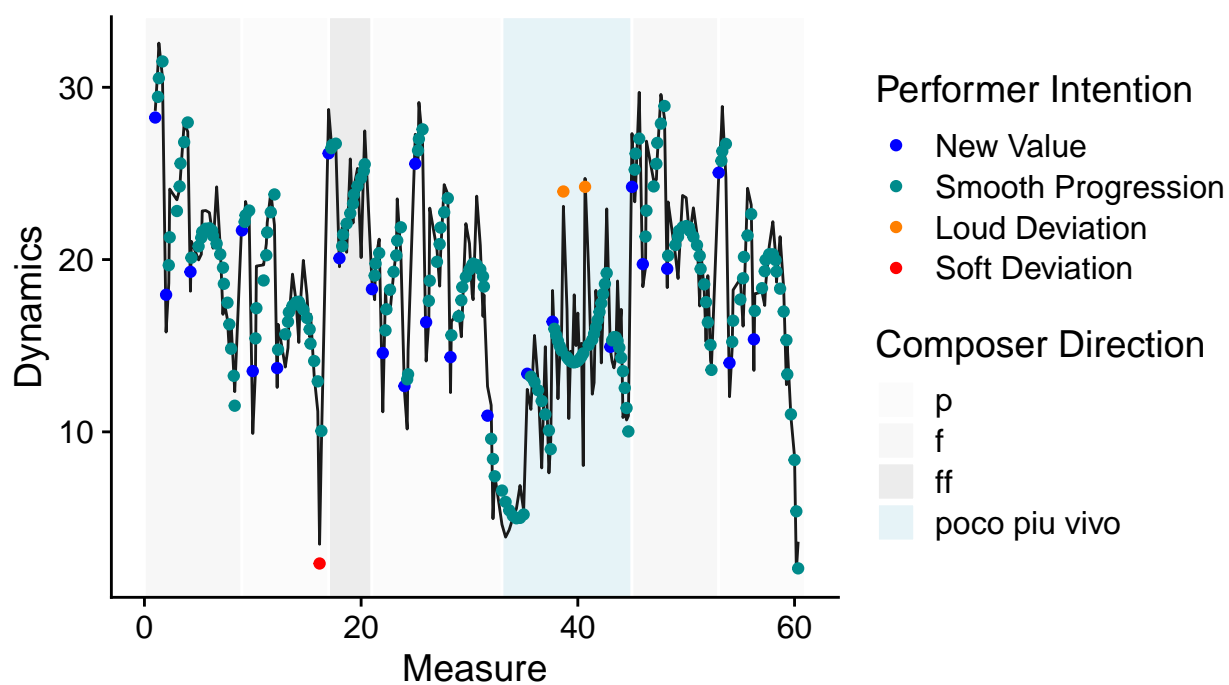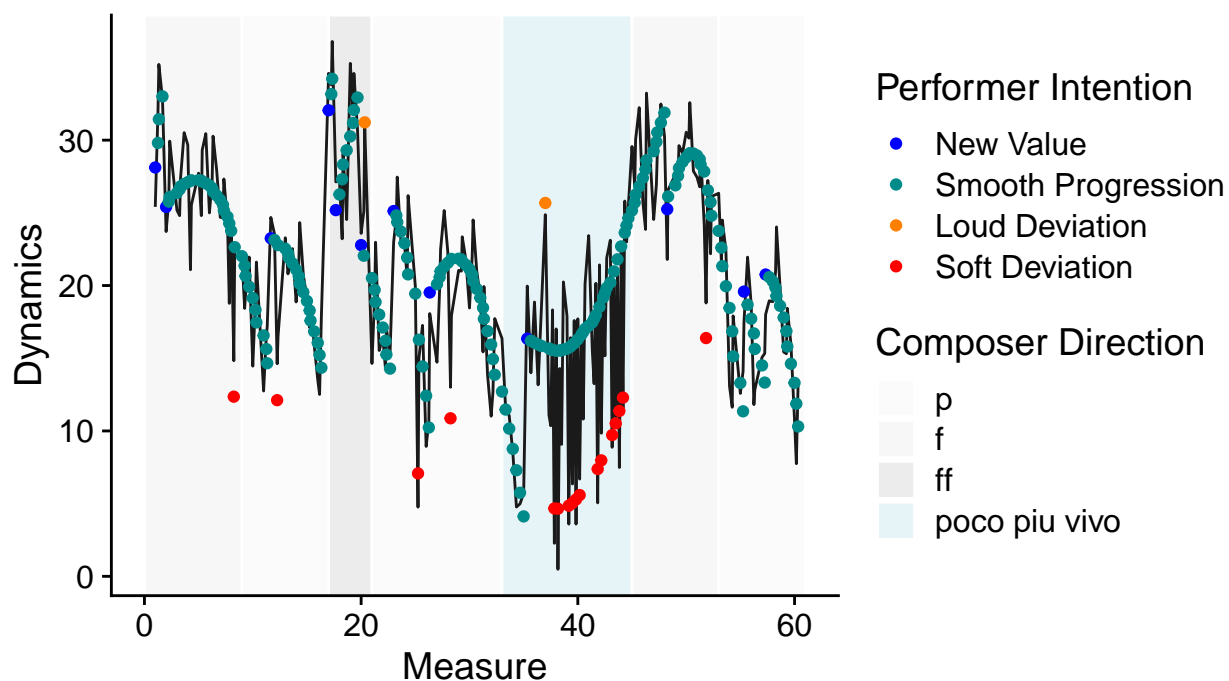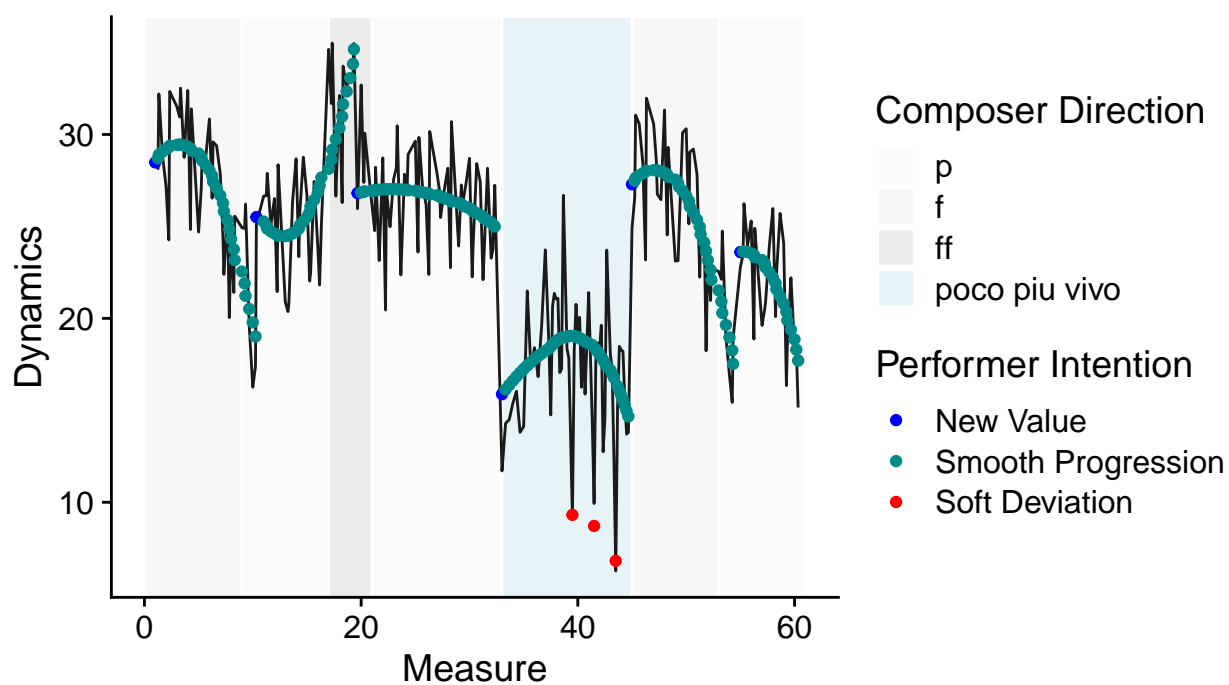
References

Akashi, H. & Kumamoto, H. (1977), 'Random sampling approach to state estimation in switching environments', *Automatica* **13**(4), 429–434.

CHARM (2009), *Centre for the History and Analysis of Recorded Music.*
URL: *http://www.charm.rhul.ac.uk/about/about.html*

Durbin, J. & Koopman, S. J. (2012), *Time Series Analysis by State Space Methods: Second Edition.*, Oxford University Press, Oxford.
URL: *http://grail.eblib.com.au/patron/FullRecord.aspx?p=991963*

Earis, A. (2007), 'An Algorithm to Extract Expressive Timing and Dynamics from Piano Recordings', *Musicae Scientiae* **11**(2), 155–182.
URL: *http://journals.sagepub.com/doi/10.1177/102986490701100202*

Fearnhead, P. & Clifford, P. (2003), 'On-line inference for hidden Markov models via particle filters', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(4), 887–899.
URL: *http://doi.wiley.com/10.1111/1467-9868.00421*

Gu, Y. & Raphael, C. (2012), 'MODELING PIANO INTERPRETATION USING SWITCHING KALMAN FILTER', *13th International Society for Music Information Retrieval Conference* .

Kalman, R. (1960), 'A New Approach to Linear Filtering and Prediction Problems1', *Journal of Basic Engineering* p. 12.

McDonald, D. J. (2020), *dpf: Discrete particle filtering.*
URL: *https://github.com/dajmcdon/dpf*

McDonald, D. J., McBride, M., Gu, Y. & Raphael, C. (2019), 'Markov-switching State Space Models for Uncovering Musical Interpretation', *arXiv:1907.06244 [stat]* .
URL: *http://arxiv.org/abs/1907.06244*

Nash, J. C. (2019), *optimr: A Replacement and Extension of the 'optim' Function.*
   URL: *https://CRAN.R-project.org/package=optimr*

Nelder, J. A. & Mead, R. (1965), 'A Simplex Method for Function Minimization', *The Computer Journal* **7**(4), 308–313.

Pandora (2020), 'Music Genome Project'.
   URL: *https://www.pandora.com/about/mgp*

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
   URL: *https://www.R-project.org/*

Rauch, H. E., Striebel, C. & Tung, F. (1965), 'Maximum likelihood estimates of linear dynamic systems', *AIAA journal* **3**(8), 1445–1450.

Spotify (2019), 'Spotify Support'.
   URL: *https://support.spotify.com/us/article/made-for-you-playlists/*

Tugnait, J. A. (1982), 'Detection and estimation for abruptly changing systems', *Automatica* **18**(5), 607–615.