TSAM: A TOOL FOR ANALYZING, MODELING, AND MAPPING THE TIMBRE OF SOUND SYNTHESIZERS

Stefano Fasciani
Faculty of Engineering and Information Sciences
University of Wollongong in Dubai
stefanofasciani@stefanofasciani.com

ABSTRACT
Synthesis algorithms often have a large number of adjustable parameters that determine the generated sound and its resultant psychoacoustic features. The relationship between parameters and timbre is important for end users, but it is generally unknown, complex, and difficult to analytically derive. In this paper we introduce a strategy for the analysis of the sonic response of synthesizers subject to the variation of an arbitrary set of parameters. We use an extensive set of sound descriptors which are ranked using a novel metric based on statistical analysis. This enables the study of how changes to a synthesis parameter affect timbral descriptors, and provides a multidimensional model for the mapping of the synthesis control through specific timbre spaces. The analysis, modeling and mapping are integrated in the Timbre Space Analyzer & Mapper (TSAM) tool, which enables further investigation into synthesis sonic response and on perceptually related sonic interactions.

1. INTRODUCTION
The timbre generated by a sound synthesis algorithm depends on the values assigned to the variable parameters, typically user configurable. Regardless of the synthesis method, the relationship between control and perceptual features of the resultant sound is generally weak [1] and difficult to determine. Modern synthesis algorithms present a wide timbre range and a high dimensional control space. The timbre, which is central in modern sonic arts, has high dimensionality as well [2] and a blurry scientific definition [3]. For designers of sonic interactive systems and of musical instruments, knowing the parameter-to-timbre relationship supports the implementation of the intended sonic response. For sound designers and performers this knowledge eases the development of control intimacy [4]. Also, this insight can help in improving the expressivity of musical instruments by reducing the control dimensionality while broadening the timbral response. The heuristic estimation of the parameter-to-timbre causality is workable, but is subjective and inaccurate. This task is challenging due to nonlinearities and correlations in the synthesis process, especially when a large set of variable parameters are involved.

We address this issue by proposing a systematic and generic method to analyze the timbre in relation to the synthesis variables. The collected data is then processed by computing a quality metric for each sound descriptor, composed of four weighted components, each representing a specific statistical characteristic. Additionally, quality metrics for synthesis parameters are provided as well. This information can be used in designing the mapping of musical gestures to the synthesis control, providing a tighter causal link with the timbral response of the system. The tool we present here, the Timbre Space Analyzer & Mapper (TSAM), integrates these functionalities and supports implementation of few-to-many lossless mappings [5], through an intermediate timbre-related layer [6]. The tool, after analyzing the sonic response of the synthesizer, computes a reduced timbre-to-parameter model, which supports real-time interaction with the sound synthesizer. In particular, we integrate an extension of the modeling and mapping strategy we introduced in [7], highlighting the enhancement achieved when considering the quality metric for selecting the descriptor for mapping purposes.

The TSAM is a flexible tool, exposing internal computation settings and options on a Graphical User Interface (GUI), which supports a range of applications and aims. The perceptual characteristics of synthesis method can be studied, characterized, and compared numerically or graphically. The relationship between timbre, spectrum and different musical scales can be investigated [8]. Different mapping approaches for musical instrument can be explored and compared. The rest of this paper is organized as follows. In Section 2 we describe the synthesis analysis procedure and present the quality metric for descriptors and parameters. Section 3 provides a summary of the timbre space mapping strategy. The TSAM implementation is detailed in Section 4. Finally, Section 5 concludes with discussion and future works.

2. TIMBRE RESPONSE ANALYSIS
Understanding the sonic variation resulting by tweaking parameters is common when getting familiar with a sound synthesizer. Different users may have distinct in-
Sound designers aim at synthesizer configurations generating the their desired sound, whereas performers and instrument builders look at a mapping that yields sonic expressivity. Synthesizers generally feature a large number of controllable parameters, representing the synthesis algorithm variables. In analog synthesizers, each parameter can theoretically assume an infinite number of values, while in digital (or software) synthesizer we have more than 4 billion possible values if considering single-precision implementations (32 bit). Synthesizers interfaced using the MIDI protocol allow only up to 128 distinct values per parameter (7 bit), despite the resolution of the internal circuitry. However with only three MIDI controlled parameters we have more than 2 million \(2^{31}\) different parameter permutations or unique synthesis states. This combinatorial explosion limits the feasibility of a comprehensive analysis of the all timbre resultant from each of these states.

Limiting the dimensionality of the parameter space allows coping with the large number of synthesis states to analyze, having only a few variable parameters and fixing the remaining to specific values. In this case the timbre analysis is limited to a subset of the entire parameter space, which is a scenario equivalent to users tweaking only a few parameters of a synthesis configuration (or preset). To further reduce the number of states to analyze we use the principle of spatial locality: states close in the parameter space generate similar timbres. Therefore we can sample the parameter space with a larger step size, and eventually interpolate at a later stage. This principle is generally true if we exclude synthesis algorithms featuring stochastic components, and parameters with strong nonlinearities (e.g. binary switches). Generally, the opposite of this principle does not hold. Proximity in the timbre space does not necessarily imply similar parameter configuration. The TSAM itself can be used to verify these principles. A further reduction can be achieved limiting the individual range of interest of each parameter.

Given \(k\) variable synthesis parameter, the synthesis state space \(I\) (set of unique parameter permutations) is given by the Equations (1)-(3) [9].

\[
I = [i_1, i_2, ..., i_n] \quad (1)
\]

\[
i = [i_1, i_2, ..., i_k] \quad (2)
\]

\[
n = \prod_{j=1}^{k} \frac{\max(i_j) - \min(i_j)}{\text{step}(i_j)} \quad (3)
\]

Each synthesis state is represented with a vector \(i\) with dimensionality \(k\), as in Equation (2), while \(n\), the number of vectors in \(I\), depends on the individual range and step size of the \(k\) parameters, as in (3). \(I\) is the synthesis state space we consider for the timbre analysis, presenting dimensionality \(k\) and cardinality \(n\).

### 2.1 Descriptors Set and Computation

For each state \(i\) of the sound synthesizer we compute a set of audio descriptors, that we indicate with \(d\), representing the timbral descriptors of the resulting synthetic sound. A large set of low-level computational descriptors, including eventual redundancies, is essential for the detailed timbre analysis we require in this context. A few higher-level timbre descriptors (e.g. brightness, noisiness, coloration), often subjective and language dependent semantic [10], are suitable to discriminate sounds with major timbral differences, but in this context they fail to capture the subtle sonic nuances determined by small variations of the synthesis parameters.

A posterior descriptor selection is possible considering the quality metric we present in this paper. The method is independent of the specific descriptors set. In the TSAM we use the CUIDADO features set [11] implemented in the IRCAM descriptors object for Max/MSP. The set includes spectral and perceptual features listed in Table 1. It includes 24 scalar and 7 vectorial descriptors, as specified in the dimensionality column, resulting in a dimensionality \(q\) of \(d\) equal to 108, as in (4). Some of the scalar descriptors in the set are closely related to traditional timbre labels (e.g. spectral centroid to brightness).

\[
d = [d_1, d_2, ..., d_9] \quad (4)
\]

<table>
<thead>
<tr>
<th>Descriptor Name</th>
<th>Dimensionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Energy</td>
<td>1</td>
</tr>
<tr>
<td>Signal Zero Crossing Rate</td>
<td>1</td>
</tr>
<tr>
<td>Spectral Centroid</td>
<td>1</td>
</tr>
<tr>
<td>Spectral Crest</td>
<td>4</td>
</tr>
<tr>
<td>Spectral Decrease</td>
<td>1</td>
</tr>
<tr>
<td>Spectral Flatness</td>
<td>4</td>
</tr>
<tr>
<td>Spectral Kurtosis</td>
<td>1</td>
</tr>
<tr>
<td>Spectral Rolloff</td>
<td>1</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>1</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>1</td>
</tr>
<tr>
<td>Spectral Spread</td>
<td>1</td>
</tr>
<tr>
<td>Spectral Variation</td>
<td>1</td>
</tr>
<tr>
<td>Perceptual Odd To Even Ratio</td>
<td>1</td>
</tr>
<tr>
<td>Perceptual Spectral Centroid</td>
<td>1</td>
</tr>
<tr>
<td>Perceptual Spectral Decrease</td>
<td>1</td>
</tr>
<tr>
<td>Perceptual Spectral Deviation</td>
<td>1</td>
</tr>
<tr>
<td>Perceptual Spectral Kurtosis</td>
<td>1</td>
</tr>
<tr>
<td>Perceptual Spectral Rolloff</td>
<td>1</td>
</tr>
<tr>
<td>Perceptual Spectral Skewness</td>
<td>1</td>
</tr>
<tr>
<td>Perceptual Spectral Slope</td>
<td>1</td>
</tr>
<tr>
<td>Perceptual Spectral Spread</td>
<td>1</td>
</tr>
<tr>
<td>Perceptual Spectral Variation</td>
<td>1</td>
</tr>
<tr>
<td>Perceptual Tristimulus</td>
<td>3</td>
</tr>
<tr>
<td>Sharpness</td>
<td>1</td>
</tr>
<tr>
<td>Spread</td>
<td>1</td>
</tr>
<tr>
<td>Noise Energy</td>
<td>1</td>
</tr>
<tr>
<td>Noisiness</td>
<td>1</td>
</tr>
<tr>
<td>Chroma</td>
<td>12</td>
</tr>
<tr>
<td>MFCC</td>
<td>13</td>
</tr>
<tr>
<td>Relative Specific Loudness</td>
<td>24</td>
</tr>
<tr>
<td>Perceptual Model</td>
<td>24</td>
</tr>
</tbody>
</table>

**Table 1.** List of descriptors used in the TSAM.

The descriptors listed above are computed on a short temporal window, typically in the range 2 ms to 200 ms. They provide an instantaneous sonic representation sufficient to characterize only absolutely periodic sounds. In synthesis states we may observe and hear low rate timbre
variations, spanning beyond the largest temporal window we consider for the descriptors. Hence an appropriate characterization of the timbre requires computation and merges of descriptors computed from multiple short time windows. We propose two analysis modes named ‘sustain’ and ‘envelope’ mode. In the first, given a synthesis state \( i \), we compute \( m \) descriptor vectors and we combine these taking their mean and optionally their range, as in Equation (5), doubling the dimensionality of the descriptor set. The second approach simply concatenates the \( m \) descriptor vectors into a single vector, as in Equation (6), increasing the dimensionality by \( m \) times.

\[
i \leftrightarrow d = \left[ \frac{\text{mean}(d_1, d_2, ..., d_m)}{\max(d_1, ..., d_m) - \min(d_1, ..., d_m)} \right] (5)
\]

\[
i \leftrightarrow d = \left[ \begin{array}{c} d_1 \\ \vdots \\ d_m \end{array} \right] (6)
\]

Considering the synthesis as an binary process, and the sound generated as almost periodic, the first approach provides a sufficient approximation of the timbre. When the synthesis produces dynamic timbres, such as texture-like sounds, or when ADSR envelopes are applied to amplitude and other parameters, the second approach is preferred. However also in presence of ADSR envelopes, we can still use the first approach, analyzing only the sustain phase of the synthesis, intentionally discarding the attack, decay and release phases, or because these do not significantly change within the parameter space \( I \) we analyze.

The concatenation of short-term static descriptors to analyze timbral dynamics is a simplification with respect to the use of dynamic descriptors computed on longer temporal windows. However this approach reduces the time needed to execute the timbre analysis and allows users to change the merging mode from ‘sustain’ to ‘envelope’ and vice versa without repeating the analysis.

In the TSAM implementation, presented in Section 4, the computation of the descriptors is completely automated. Users are only required to identify the \( k \) variable parameters of the synthesizer, their range, step size, number of descriptor vectors per state \( m \), and analysis mode. The tool computes \( I \) and drives the synthesizer with one \( i \) at a time, computing and storing \( m \) vectors \( d \). For analysis in envelope mode, the tool also manages the triggering of the synthesizer at every \( i \). Users can further specify the temporal unfolding of the analysis, selecting only a subset of the ADSR envelope. Advanced options related to the descriptor computation, such as window size, hop size, sampling rate, are exposed as well.

### 2.2 Descriptor Quality Metric

The quality metric we compute for each descriptor is aimed at capturing the four characteristics listed below.

- **Noisiness**: deviation of the descriptor from its mean given a synthesis state \( i \).
- **Variance**: spread of descriptor value across the synthesis state space \( I \).
- **Independence**: uniqueness of the descriptor variation pattern across the synthesis state space \( I \).
- **Correlation**: coherence of the descriptor variation with synthesis parameters across the synthesis state space \( I \).

Ideally, a descriptor representative of \( I \) should present low noisiness, high variance, high independence, and high correlation. High noisiness indicates that a particular descriptor and the associated timbral characteristic also varies when synthesis parameters are fixed, and therefore its eventual variance across \( I \) may be not significant. A descriptor with low variance reveals that the related timbral characteristic does not change significantly when varying the synthesis parameters. Descriptors varying with a similar trend are redundant, and thus less significant, when computing a dimensionality-reduced timbre space modeling \( I \), instead those more independent carry a larger amount of information. Descriptors can also be highly independent when varying randomly across \( I \). We address this by also including the correlation between descriptor and parameters in the metric, as we expect representative descriptors to change accordingly to one or more synthesis parameter.

For each descriptor, we compute the noisiness \( N_{x,i} \) from the \( m \) descriptor vectors in synthesis state \( i \) before these are merged, as per Equations (5) and (6). The subscript \( x \) is the index identifying the descriptor across the set of \( q \) computed in the TSAM. For ‘sustain’ mode, we measure the deviation of the descriptor \( x \) in the state \( i \) using the Relative Mean Absolute Difference (RMD), as in Equation (7). The RMD is a scale invariant measure of statistical dispersion, hence allows the comparison of heterogeneous descriptors. For ‘envelope’ mode, \( N_{x,i} \) is estimated as the zero crossing rate, as in Equation (8), of the forward second order finite difference (discrete approximation of the second order derivative) of the series of \( m \) descriptors, as in (9). This represents the rate at which a descriptor inverts its trend (from increasing to decreasing and vice versa) in the analyzed envelope. Noisy descriptors invert their trends at higher rates.

\[
N_{x,i} = \sum_{j=1}^{m} \sum_{k=1}^{m-1} |d_{x,j} - d_{x,k}| / \sum_{j=1}^{m} d_{x,j} (m - 1) (7)
\]

\[
N_{x,i} = \frac{1}{m-2} \sum_{j=2}^{m} \{ \Delta^2(d_{x,j}) \Delta^2(d_{x,j-1}) < 0 \} (8)
\]

\[
\Delta^2(d_{x,j}) = \sum_{k=0}^{2} \left( \frac{2}{k} \right) (-1)^{2-k} d_{x,j+k} (9)
\]

In Equations (7)-(9), \( d_{x,j} \) represents the \( x \)-th descriptor in the set of \( q \), from the \( j \)-th vector \( d \) out of the \( m \) computed for each state \( i \). The indicator function \( \mathbb{I} \) is equal to 1 if its argument is true, 0 otherwise. \( \Delta^2( \cdot ) \) is the forward second order finite difference function. The overall noisiness of each descriptor \( N_x \) is computed by taking the average over the set of synthesis unique states \( I \) we analyze.
Variance, independence, and correlation are computed across $\mathbf{I}$, after the $m$ descriptors are merged as in (5)-(6). The same method is used for both ‘sustain’ and ‘envelope’ modes. The variance $V_x$ is computed as the RMD over the $n$ synthesis states $i$. We use the same expression as in (7), replacing $m$ with $n$, but in this case $d_{x,j}$ is the $x$-th descriptor in the set of $q$, from the $j$-th vector $\mathbf{d}$ out of the $n$ we compute across $\mathbf{I}$.

We assume that descriptors are independent if poorly correlated, therefore we compute $I_x$ taking the complement of the averaged absolute value of the correlation coefficient between the descriptor $x$ and the other $q$-$i$ descriptors over $\mathbf{I}$, as in Equation (10). Both positive and negative correlations indicate dependence, therefore we take the absolute value of the correlation coefficient $|\text{corr}(\cdot)|$. We subtract 1 from the summation to remove the correlation coefficient of the descriptor with itself, when $j=x$. Finally, the correlation $C_x$ between descriptors and parameters is computed taking the average correlation coefficient between the $x$-th descriptor and the $k$ variable synthesis parameter, as in Equation (11).

$$I_x = 1 - \frac{1}{q-1} \left[ \left( \sum_{j=1}^{q} |\text{corr}(d_{x,p}, d_{j,l})| \right) - 1 \right]$$  \hspace{1cm} (10)

$$C_x = \frac{1}{k} \sum_{j=1}^{k} |\text{corr}(d_{x,p}, i_{j,l})|$$  \hspace{1cm} (11)

In (10) and (11) with $d_{x,1}$ we represent the vector containing the $n$ values of the $x$-th descriptor computed over the synthesis state space $\mathbf{I}$, while $i_{j,1}$ represents the vector containing the $n$ values of the $x$-th synthesis parameter over $\mathbf{I}$. Note that according to (5) and (6) each descriptor may contribute with more than one component in each vector $\mathbf{d}$. In particular, for ‘sustain’ mode we have two components per descriptor if the range is included in the analysis, whereas for the ‘envelope’ mode we have $m$ components per descriptor. Therefore we compute multiple $V_x$, $I_x$ and $C_x$ per each $x$-th descriptor, and use their average in the quality metric we introduce next.

The quality metric $S_x$ of each descriptor is computed from the individual noisiness, variance, independence, and correlation as in Equation (12). The noisiness, being an undesirable feature, lowers the value of $S_x$. The four components are combined using individual weights $w$. 

$$S_x = w_p V_x + w_I I_x + w_C C_x - w_N N_x$$  \hspace{1cm} (12)

The selection of the $w$ values depends on the aim and context of the timbre analysis, and also on individual preferences. For instance, when analyzing a synthesizer configuration with a texture-like timbre, we expect considerable sonic variation within each synthesis state $i$, therefore the noisiness has no significance and $w_N$ should be close to zero. If the purpose of the analysis is the sole study of the synthesizer timbre through the descriptors, their independence has little relevance. Instead when descriptors are used for mapping purposes, as in Section 3, the independence has a higher significance. In the TSAM, the default values of the weights are 0.33 for variance, independence and correlation, and 0.66 for noisiness. Users can change these in the unitary range. The four components of the quality metric have different ranges. $I_x$ and $C_x$ span between [0,1], while $N_x$ and $V_x$ can be zero but do not have a theoretical maximum. In the TSAM we include the option to normalize these to the unitary range, easing the balancing through individual weights. However when comparing the quality metrics $S_x$ across different synthesizers, or between different state spaces $\mathbf{I}$ of the same synthesizer, normalization should not be used. In the TSAM we also rank also the $k$ synthesis parameter by their average correlation with the $q$ descriptors, computed as in (11) but replacing $k$ with $q$ and taking $x$ as the summation index. Furthermore for each parameter, the TSAM displays the two descriptors with associated highest and lowest correlation, and vice versa.

3. TIMBRE SPACE MODELING AND MAPPING

Audio descriptors have been extensively used for visualization, measurement, classification, and recognition of sounds. Works proposing the timbre as a control structure for sound synthesis [12] or for interactive sonic systems have recently proliferated [13]–[24]. These allow for explicit control of psychoacoustic characteristics of the generated sound, hiding synthesis parameters from users, simplifying the user interaction, facilitating the search for specific timbres, and enhancing the expressivity of the system. Similar benefits are provided by synthesis methods using a timbre representation derived by a prior analysis stage of the target sound [25], [26]. A model relating parameters to sonic response of the sound synthesizer is necessary to implement explicit timbre control. Our generic approach, introduced in [7] and extended here, derives a model from the prior analysis stage, and therefore it is independent of the specific synthesis method and implementation.

The generative mapping is based on unsupervised machine learning techniques, and it provides a low dimensional and perceptually related synthesis control. The mapping maximizes the breadth of the explorable sonic space covered by the synthesis space $\mathbf{I}$, and minimizes possible timbre losses due to the reduced dimensionality of the control space (i.e. few-to-many mapping). The timbre response analysis described in the previous section returns a synthesis space $\mathbf{I}$, with dimensionality $k$, and a descriptor space $\mathbf{D}$, with dimensionality $q$. Both spaces present $n$ entries $i$ and $d$, which are pairwise associated, representing a basic model relating parameters and timbre. Hence we can explicitly express a timbre through the $q$ descriptors (e.g. mapped on a large bank of faders), find the closest entry in $\mathbf{D}$, and drive of the synthesizer with the associated parameter set $i$. Such control is affected by several drawbacks: the high dimensionality of the timbre-based control, with $q$ generally much greater than $k$; the lack of accuracy due to the large parameter step size we use in the analysis stage (3); entries in the timbre space $\mathbf{D}$...
are not evenly distributed as in $\mathbf{I}$, hence regions of $\mathbf{D}$ with low density determine a poor system response.

The real dimensionality of $\mathbf{D}$ is usually much less than $q$. Generally the data of interest lies on an embedded non-linear manifold within the $q$-dimensional space. Therefore we reduce the dimensionality of $\mathbf{D}$, using Isomap, down to two or three dimensions, which are easy to map to general-purpose controllers with low cognitive complexity. In the TSAM users can explore the application of 34 different dimensionality reduction methods [27].

Before reducing the dimensionality of $\mathbf{D}$, we use the quality metric $S_2$ to discard those descriptors with a low score. Particularly noisy or poorly correlated descriptors present a large variance that have a significant impact in the dimensionality reduction stage, but this would not be representative of the parameter-to-timbre relationship, corrupting the timbre space mapping. The selection of descriptors based on the quality metric determines improvements in accuracy and usability against our previous approach. Alternatively, users can bypass the dimensionality reduction stage, and explicitly specify the two or three descriptors composing the low dimensional timbre space we use for the mapping to synthesis parameters.

To address the issue of the possible unresponsiveness of the timbre space due to arbitrary distribution in $\mathbf{D}$ we apply an iterative algorithm based on the Voronoi tessellation, derived from [28], that redistribute the $n$ entries $\mathbf{d}$ into an uniformly distributed square or cube, while preserving the local neighborhood relationships (homomorphic transformation). The inverse of this transformation represent the required mapping to project a generic multidimensional control vector $\mathbf{C}$ onto the specified timbre space. Hence we use an Artificial Neural Network (ANN) to learn a function $m(\cdot)$ approximating the inverse of the redistribution process. We use $m(\cdot)$ to project the generic multidimensional control vector $\mathbf{e}$ onto the dimensionally reduced timbre space $\mathbf{D}^t$. The ANN includes a single hidden layer and therefore can be trained efficiently using a non-iterative algorithm [29].

In Figure 1 we show an example of a highly clustered timbre space reduced to three dimensions, and its transformation to a uniform cube. The side arrows identify the two stages of the mapping computation. In the TSAM we provide also an alternative mapping, skipping the ANN and computing the synthesis parameters directly from the uniformly distributed timbre space.

In the final stage of the mapping we compute the parameters to interact with the sound synthesizer. We use $\mathbf{d}^*$ to represent a descriptor vector in the dimensionality reduced timbre space $\mathbf{D}^t$. Driving the synthesis with the parameters $\mathbf{i}$ associated with the $\mathbf{d}^*$ closer to $m(\mathbf{e})$ may lead to discontinuities, that in turn may generate glitches in the sonic output. These are due to the coarse parameter step size used in the analysis stage, and due to the not one-to-one relationship between parameters and sound. Two synthesis states $\mathbf{i}$, far apart in the synthesis state space $\mathbf{I}$, may be associated identical or similar descriptor vectors $\mathbf{d}$, hence close in $\mathbf{D}$. The latter is an implicit drawback of any methods for controlling sound synthesis from any representation of the generated signal.

We address these issues computing the synthesis parameter by spatial interpolation, including only entries of $\mathbf{D}^t$ from the neighborhood the current state $\mathbf{i}$. The set of parameters driving the synthesizer $\mathbf{i}_{c\text{tri}}$ is computed by Inverse Distance Weighting (IDW) as in Equations (12) and (13), where $\| \cdot \|$ represent the Euclidean distance.

$$\mathbf{i}_{c\text{tri}} = \frac{\sum_{j=1}^{N} \mathbf{q}_j(m(\mathbf{e})) \cdot \mathbf{i}_j}{\sum_{j=1}^{N} \mathbf{q}_j(m(\mathbf{e}))} \quad (12)$$

$$\mathbf{q}_j(m(\mathbf{e})) = \frac{1}{\| m(\mathbf{e}) - \mathbf{d}_j \|^p} \quad (13)$$

In (12) and (13) $N$ represents the total number of points considered in the interpolation, and the $\mathbf{i}_j$ in (12) are those pairwise associated with the $\mathbf{d}_j$ in (13). In the TSAM instead of using the $N$ closest point $\mathbf{d}_j$ in $\mathbf{D}^t$, we select those $\mathbf{d}_j$ that limit the maximum variation of $\mathbf{i}_{c\text{tri}}$ between two consecutive iterations, that is the set of $\mathbf{d}_j$ associated with the $\mathbf{i}_j$ close to the current $\mathbf{i}_{c\text{tri}}$ (within a user-defined distance). In Figure 2 we show an example of this interpolation points selection, where the green entries are the $\mathbf{d}_j$ related to $\mathbf{i}_j$ close to the current $\mathbf{i}_{c\text{tri}}$, which is in turn associated with the yellow one in figure.
The set of $d^j_i$ used for IDW interpolation may include entries distant from $m(e)$, but these will poorly contribute in (12). In the IDW, $p$ represents the power parameter, which determines the influence of each point based on the distance. This value should be larger than the dimensionality of the reduced timbre space $D^d$, and increasing $p$ closer points has larger weight. In the TSAM, the $i_{crit}$ maximum instantaneous distance and interpolation power parameter $p$, are among the options exposed to users to tune in real time the timbre mapping response. The TSAM provides interactive timbre space visualizations, such as those in Figure 1 and 2.

Figure 2. Detail of a timbre space reduced to three dimensions. The green entries are those used in the interpolation to compute the synthesis parameter, because close to the yellow current entry in the synthesis state space.

4. IMPLEMENTATION AND USAGE

The TSAM$^1$ is an open-source software implemented in Max/MSP using FTM extension$^2$ [30], supported by a background engine written and compiled in MATLAB. The analysis of the synthesis timbre, the real-time timbre space mapping and the visualizations are computed in Max/MSP, whereas the background engine computes the descriptor quality and the timbre space mapping (dimensionality reduction, redistribution, ANN training), taking as input the outcome of the analysis stage. The two components of the system communicate via Open Sound Control (OSC) protocol and large matrices are exchanged using files. The TSAM can host software synthesizer developed using Steinberg’s Virtual Studio Technology (VST). It acts as a wrapper for VST synth, providing a fully integrated environment. The TSAM allows full control of all parameters for analysis and mapping purposes. It captures the synthesized signal for descriptor computation and playback, and manages the global state of the synthesizer when saving and restoring presets. In Figure 3 there is a screenshot of the main TSAM GUI. This exposes a large number of options for further exploration of the mapping method we propose, and also for customizing analysis, mapping computation, real-time control, and visualization. Default settings are provided for basic use. Users can load a VST synth and select up to 10 variable parameters, their range, analysis step size, and the number of vectors $m$ per state $i$. Advanced analysis options include digital signal processing settings and analysis timing with respect to the synthesis triggering (note-on and note-off messages). The TSAM estimates and shows the total analysis time, and users may opt to reduce the parameter step sizes, in (3), when this is excessive. Thereafter the analysis is carried out automatically. In Section 2 we discussed two analysis modes, ‘sustain’ and ‘envelope’ respectively. These, besides the automatic mode, can also be carried out manually. Users arbitrarily tune the synthesizer to a specific state $i$, and request for the descriptor analysis of the related sonic response (both modes are supported). Furthermore we included the interactive ‘sustain’ analysis mode [7] where descriptor vectors $d$ are computed while users vary in the MIDI mapped synthesis parameters in real-time, dynamically generating a stream of $i$. The latter analysis mode does not guarantee to observe an identical number of descriptor vectors $d$ per state $i$, hence the noisiness in the quality metric result may be inconsistent.

When the analysis stage is completed, users can request the computation of the descriptor quality metric, which is visualized in the TSAM as shown in Figure 4. In the descriptors page, users can also specify the weights of Equation (12), enable the normalization of its components, find and rank the descriptors by highest score, observe the synthesis parameter ranking, and find the highest and lowest correlation between each parameter and descriptor. Furthermore, users can specify which subset of the 108 descriptors will be used for mapping purposes.

Options for the timbre space mapping computation include the dimensionality of the map, selection of the dimensionality reduction technique and the ANN activation function. The mapping can be tuned at runtime using the settings discussed in Section 3. The timbre analysis, quality metric, and mapping are saved into files that can be individually recalled through the TSAM presets.

5. DISCUSSION AND FUTURE WORK

We presented a generic tool that integrates functionalities to study and map the timbre of sound synthesizers. Preliminary studies demonstrated that the adoption of large sets of descriptors, and their selection based on the novel quality metric, improves the accuracy of the timbre-based interaction. The TSAM can be used for the study of the sonic response of synthesizers, for an explicit control of timbral character, or for a reduction of the synthesis control space, exposing only a few perceptually relevant control dimensions. Previous user studies on a system with a similar mapping approach demonstrated that synthesis parameters become transparent to users [31], which are exclusively focused on the timbral interaction. Future works include user studies with the TSAM to evaluate the

---

1 http://stefanofasciani.com/tsam.html
2 http://ftm.ircam.fr/
effectiveness of the timbre-based mapping, comparing it against traditional and alternative approaches to sound synthesis interaction, in performing and sound design scenarios. Moreover we will investigate the relevance of different descriptor categories for a more perceptually related sonic control.

**Figure 3.** TSAM main page, including options for analysis, mapping computation, real-time control, and visualization.

**Figure 4.** TSAM descriptor page, providing an insight into the timbre response and parameter relationship of the synth.
6. REFERENCES


