

Timbre remapping through a regression-tree technique

Dan Stowell and Mark D. Plumbley

Centre for Digital Music, Queen Mary University of London, UK

dan.stowell@elec.qmul.ac.uk

ABSTRACT

We consider the task of inferring associations between two differently-distributed and unlabelled sets of timbre data. This arises in applications such as concatenative synthesis/audio mosaicing in which one audio recording is used to control sound synthesis through concatenating fragments of an unrelated source recording. Timbre is a multidimensional attribute with interactions between dimensions, so it is non-trivial to design a search process which makes best use of the timbral variety available in the source recording. We must be able to map from control signals whose timbre features have different distributions from the source material, yet labelling large collections of timbral sounds is often impractical, so we seek an unsupervised technique which can infer relationships between distributions. We present a regression tree technique which learns associations between two unlabelled multidimensional distributions, and apply the technique to a simple timbral concatenative synthesis system. We demonstrate numerically that the mapping makes better use of the source material than a nearest-neighbour search.

1. INTRODUCTION

This paper aims to improve musical expression by audio-based control of timbre. There are various applications in which the timbral analysis of a sound is used to control a system whose output is sound of a different type, such as concatenative synthesis [1][2][3], query-by-example [4] or adaptive digital audio effects [5]. In such cases there are two different timbral distributions to consider – that of the controlling sound, and that of the audio output – and we wish to be able to map from one to the other. Often the two distributions are quite different, since for example we may wish to map from vocal sounds on to timbres which cannot be produced vocally, so the issue of mapping is non-trivial.

In this paper we argue that mapping through standard techniques such as nearest-neighbour search may be insufficient, and we present a new nonparametric technique based on regression trees which accounts for the differences between timbral distributions. We then apply the technique in a simplified timbral concatenative synthesis system, and demonstrate numerically that the mapping

makes better use of the source material than a nearest-neighbour search.

1.1 Timbre trajectories: absolute or relative?

In order to map the timbre trajectory of a sound (its evolution over time) onto another, we will require some timbre analysis of the signal. An issue that affects our choice of search strategy is whether the timbral analysis should best be treated as absolute context-independent data, or whether it should be treated as relative – for example, relative to the range of the sound source which produced it. Given a particular timbral “coordinate”, should we treat it differently if we knew that it was produced by a clarinet or by a violin? Would such information imply a difference in the expressive purpose of the sound?

The common definition of timbre describes it as that attribute which enables a listener to differentiate sounds which are equal in pitch and loudness [6]. It therefore does not demand that timbre be an absolute or context-invariant attribute of a sound. Research into music timbre perception has taken a similar stance, basing experiments on comparisons among sets of sound examples [7, 8, 9, 10]. Such studies often explain results in part through acoustic features derived from the examples, which can imply a context-independent notion of timbre inherent in the signal. However Grey [11] finds evidence for context-dependence of timbre perception in musical patterns. Lakatos [12] offers some consideration of contextual effects by investigating sets of harmonic and percussive sounds both separately and combined. He presents evidence supporting the existence of two broadly context-independent timbre dimensions but also for some degree of contextual influence on timbre judgements.

Musical applications of timbre analysis often use acoustic features taken from the signal (e.g. [13][14, Chapter 16]), implicitly treating timbre as absolute. This will certainly be appropriate in situations where the timbre data contains strong semantic “anchors” – a clear example of this occurs in human speech, where vowels are largely characterised by the absolute positions of the main resonances (formants) on the frequency scale [15]. However, the evidence of context-dependence in musical timbre suggests this may not always be the case. Consider a system which synthesises sound based on timbral examples produced by voice (e.g. [14, Part III]): the human voice is naturally constrained to its own timbre range, yet we may well wish to induce the system to produce sounds outside this range. In fact we consider this to be a basic requirement, since such ability to extend our timbral range is one of the main ap-

peals of such technologies.

1.2 Timbre lookup strategies

The most basic form of timbral search is perhaps a nearest-neighbour (NN) search [16], often using Euclidean distance. Since timbre features in general have quite different ranges, their ranges may be standardised before search, or a scale-invariant metric such as Mahalanobis distance may be used [17]. For example, Schwarz [14, Chapter 16] uses the Euclidean distance normalised over the entire database of sounds. This normalisation accounts for differences between the ranges of the features, but not for differences between the timbral range of the different sound sources included in the database. Note that timbral distance search is but one criterion used in a concatenative synthesiser such as this, which uses a constraint-satisfaction framework to combine criteria related to duration, pitch and other considerations.

Large database search systems often do not store the raw timbral co-ordinates needed for NN search, but parametrically model the timbre of a recording (e.g. using Gaussian Mixture Models) and store the model parameters [13]. Timbral search can then be performed by finding the parameter-set which maximises the likelihood of query data.

Whether search is performed by instance-based methods such as NN or model-based methods such as Gaussian Mixture Model likelihood, the difference in timbral ranges of different sound sources is often neglected, perhaps reflecting an approach to timbre as absolute rather than relative. One approach to account for this could be to standardise the mean and variance of timbre features for each type of sound source, or for each recorded audio excerpt, which would accommodate the large-scale differences. However it would fail to account properly for multidimensional interactions in the data such as the movement of one region relative to the rest of the distribution.

Rather than pursuing the idea of a normalisation scheme as a precursor to search, in this paper we develop an integrated method which automatically learns to map from one data distribution to another, assuming similarities in the orientation of the datasets in timbre space but allowing for differences in the distributions at large and small scales. Tree methods are attractive in this context because recursive partitioning provides a generic approach to dividing multidimensional distributions into regions of interest at multiple scales. We next describe the method, before applying it in a concatenative synthesis experiment.

2. AUTO-ASSOCIATIVE REGRESSION TREES

A regression tree [18, Chapter 8] is a computationally efficient nonparametric way to analyse structure in a multivariate dataset, with a continuous-valued response variable to be predicted by a set of independent variables. The core concept is to recursively partition the dataset, at each step splitting it into two subsets using a threshold on one of the independent variables (i.e. a splitting hyperplane orthogonal to one axis). The choice of split at each step is made

to minimise an “impurity” criterion for the value of the response variable in the subsets, often based on the mean squared error [18, Section 8.3]:

$$\text{impurity}(\alpha) = \sum_{i=1}^{n_\alpha} (y_i - \bar{y})^2 \quad (1)$$

where n_α is the number of data points in the subset α under consideration, and \bar{y} the mean of the sampled values of the response variable y_i for the points in α .

The original formulation of regression trees was concerned with predicting a single univariate response variable. They were subsequently extended to multivariate responses, for example by [19] using a direct multivariate extension of (1):

$$\text{impurity}(\alpha) = \sum_{i=1}^{n_\alpha} \sum_{j=1}^p (y_{ij} - \bar{y}_j)^2 \quad (2)$$

with definitions as in (1) except that the y_i (and therefore also \bar{y}) are now p -dimensional vector values, with j indexing over the dimensions.

This extension yields a framework that can learn to infer relationships between one multivariate data distribution (the independent variables) and another (the response) – hence their potential application to the inference of mapping from one set of multivariate timbre data to another. One limitation of this is that the regression is still a supervised technique, meaning that the pairwise association between items in the training datasets would need to be provided. In applications such as ours, where we might have a large database of short audio fragments from various sources, it will often be impractical to annotate the data, so we seek an unsupervised method. We will develop an existing unsupervised application of regression trees for this task.

2.1 Auto-association and multivariate splits

Questier et al. [19] apply regression trees to the task of discovering structure in unsupervised multivariate data, by equating the response variables with the independent variables, to create an *auto-associative* multivariate regression tree (AAMRT). In other words they apply a standard regression tree with the multivariate-response extension, but there is no separation between the variables used to split the dataset and the variables whose impurity is to be minimised – the independent variables are made to “predict themselves”. This is reminiscent of data-driven histogramming [20]; in the work of Questier et al. it is used for feature-selection by analysing which features are most commonly used for splitting.

There are in fact two types of multivariate extension to the standard regression tree. We have already described the *multivariate-response* extension; also the choice of splitting plane can be generalised so that it can take any orientation in the feature space rather than being aligned with one axis [21]. We refer to this as the *multivariate-splits* extension. Gama [22] shows that this extension can reduce bias in the resulting estimator. Further, it may make more

effective use of the available information if there is a limited number of datapoints: if there are N data points then there can be no more than around $\log_2 N$ splits used to reach a leaf in a balanced binary tree. This could well be fewer than the number of dimensions, meaning the information from some dimensions would be neglected. For the remapping task discussed in the next section, then, we will use a regression tree that is based on AAMRT but multivariate in both senses.

Note that the impurity measures (1) & (2) are equivalent to the sum of variances in the subsets, up to a multiplication factor which we can disregard for the purposes of minimisation. By the law of total variance (see e.g. [23, Appendix S]), minimising the total variance within the subsets is the same as maximising the variance of the centroids; therefore the impurity criterion selects the split which gives the largest difference of the centroids of the response variable in the resulting subsets. If only univariate splits are allowed then this can be optimised as given in [18, Chapter 8]. In the multivariate-splits variant, maximising the variance of the centroids is achieved simply by selecting the hyperplane perpendicular to the first principal component in the (centred) data. This multivariate-splits variant of AAMRT allows for efficient implementation since the leading principal component in a dataset can be calculated efficiently e.g. by expectation-maximisation [24].

3. CROSS-ASSOCIATION

We wish to generalise the AAMRT method to apply it to two datasets defined on the same space. A simple approach would be to combine the two datasets into one and then apply AAMRT, but this would not allow the algorithm to adapt separately to the two datasets, to account for differences in location.

Instead, we modify the algorithm so that at each step of the recursion the data coming from the two distributions are separately centred. One single principal component is then calculated from their union. The recursion therefore generates two “similar but different” trees, implementing the notion that the two datasets have similarities in structure (the orientations of the splitting planes are the same) but may have differences in location at various scales (the centroids of large or small subsets of the data are allowed to differ). This is illustrated schematically in Figure 1.

If the datasets are unequal sizes then the larger set will tend to dominate over the smaller in calculating the principal component. To eliminate this issue we weight the calculation so as to give equal emphasis to each of the datasets, equivalent to finding the principal component of the union J of weighted datasets:

$$J = \bigcup (N_Y(X - C_X), N_X(Y - C_Y)) \quad (3)$$

where X and Y represent the data (sub)sets, C_X and C_Y their centroids, and N_X and N_Y the number of points they contain.

The resulting cross-associative multivariate regression tree (XAMRT) algorithm is summarised in Figure 2. Note that we do not prune the tree [18, Chapter 3], since for the

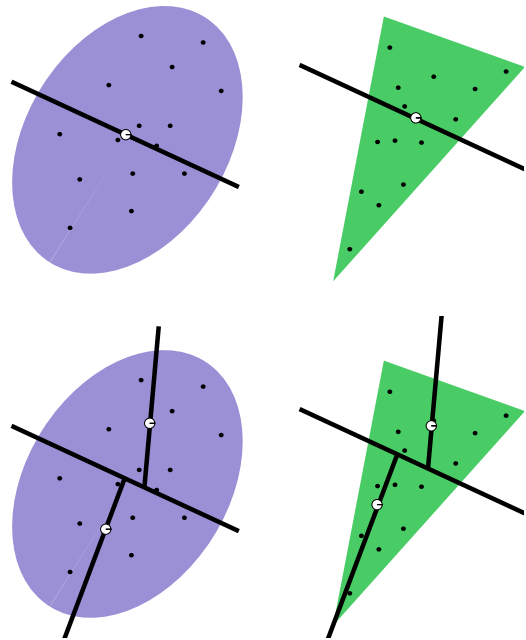


Figure 1. Schematic representation of the first two steps in the recursion. In the first step (top), the centroids of each dataset are calculated separately, and then a splitting plane with a common orientation is chosen. The second step (bottom) is the same but performed separately on each of the partitions produced in the first step.

timbral application presented here, all of the variation in the training set is useful for resynthesis.

To perform a remapping using a XAMRT data structure, one takes a data point and descends the tree, at each split centring it by subtracting C_X or C_Y as appropriate and then deciding which side of the splitting plane it falls. When the leaf node is reached, it contains two sets of training data points (a subset each of X and Y). To choose a corresponding coordinate relating to the opposite distribution, one could for example use a random datum selected from the opposite subset, or the centroid of that subset, depending on the application. (If the sizes of the datasets are similar then the leaf will often contain just one datum from each of the two distributions.)

4. TIMBRE REMAPPING EXPERIMENT

Our algorithm can be applied to timbre remapping tasks, i.e. ones in which the timbral trajectory of a sound source is used to control that of some other system. Concatenative synthesis is an example of such a task, in the case where an input sound is used as the controller for the concatenative synthesiser (also referred to as audio mosaicing). Concatenative synthesis is not the only example of timbral application of our algorithm – we are investigating application of the technique in general synthesiser control – but it presents a known system in which timbres from heterogeneous sources are used to control sound generation.

Numerical evaluation of timbre remapping quality is difficult since the perceptual quality and musical merit of the audio result have no obvious objective metric. How-

XAMRT(X, Y)

```

 $C_X \leftarrow$  centroid of  $X$ 
 $C_Y \leftarrow$  centroid of  $Y$ 
 $J \leftarrow$  result of equation (3)
 $p \leftarrow$  principal component of  $J$ 
 $X_l \leftarrow X \cap ((X - C_X) \cdot p > 0)$ 
 $X_r \leftarrow X \cap ((X - C_X) \cdot p \leq 0)$ 
 $Y_l \leftarrow Y \cap ((Y - C_Y) \cdot p > 0)$ 
 $Y_r \leftarrow Y \cap ((Y - C_Y) \cdot p \leq 0)$ 
if  $X_l$  is singular or  $Y_l$  is singular
  then  $L = [X_l, Y_l]$ 
  else  $L = \text{XAMRT}(X_l, Y_l)$ 
if  $X_r$  is singular or  $Y_r$  is singular
  then  $R = [X_r, Y_r]$ 
  else  $R = \text{XAMRT}(X_r, Y_r)$ 
return  $[L, R]$ 

```

Figure 2. The cross-associative algorithm. X and Y are the two sets of vectors between which associations will be inferred.

Description	Duration (sec)	No. of grains
Amen breakbeat	7	69
Beatboxing	93	882
Fireworks	16	163
Kitchen sounds	49	355
Thunder	8	65

Table 1. Audio excerpts used. “No. of grains” is the number of 100ms grains segmented and analysed from the audio (excluding silent frames) – see text for details.

ever, concatenative synthesis offers the opportunity for numerical evaluation by studying the statistics of usage of the different grains, as will be described in Section 4.4.

We require an experiment which will probe the timbral matching performance of our algorithm. Concatenative synthesisers typically operate not only on timbre, but use pitch and duration as well as temporal continuity constraints in their search strategy, and then modify the selected grains further to improve the match [25]. While recognising the importance of these aspects in a full concatenative synthesis system, we designed an experiment in which the role of pitch, duration and temporal continuity were minimised, by excluding such factors from grain construction/analysis/resynthesis, and also by selecting audio excerpts whose variation is primarily timbral.

We first describe the audio excerpts we used and how timbre was analysed, before describing the concatenative synthesiser and our performance metric.

4.1 Audio data

In order to focus on the timbral aspect, we selected a set of audio excerpts in which the interesting variation is primarily timbral and pitch is less relevant. The five excerpts – two musical (percussive) and three non-musical – are listed

in Table 1 and are also available online.¹ The excerpts are 44.1 kHz mono recordings.

The excerpts are quite heterogeneous, not only in sound source but also in duration (up to an order of magnitude). They each contain various amounts/types of audio event, which are not annotated. This wide variety of excerpts was selected to give a clear impression of the success of the remapping techniques at drawing timbral analogies.

4.2 Timbre features

We chose a set of 10 common acoustic timbre features: spectral power, spectral power ratio in 5 log-spaced sub-bands (50–400, 400–800, 800–1600, 1600–3200, and 3200–6400 Hz), spectral centroid, spectral 95- and 25-percentiles and zero-crossing rate (for definitions see [26]).

Analysis was performed on audio “grains”: units of fixed 100ms duration taken from the audio excerpt every 100ms (i.e. with no overlap). Each grain was analysed by segmenting into frames of 1024 samples (at 44.1 kHz sampling rate) with 50% overlap, then measuring the feature values for each frame and recording the mean value of each feature for the grain. Grains with a very low spectral power (< 0.002) were treated as silences and discarded. The timbre features of the remaining grains were normalised to zero mean and unit variance within each excerpt. Analysis was performed in SuperCollider 3.3.1 [27].

4.3 Timbral concatenative synthesiser

We designed a simple concatenative synthesiser using only timbral matching, either by a standard nearest-neighbour (NN) search or by our algorithm. Given two excerpts – one which is the source of grains to be played back, and one which is the control excerpt determining the order of playback – and the timbral metadata for the grains in the two excerpts, the synthesis procedure works as follows:

For each grain in the control excerpt, if the grain is silent (power < 0.002) then we replace it with silence. Otherwise we replace it with a grain selected from the other excerpt by performing a lookup of the timbre features – either a NN search or the XAMRT tree regression. For numerical evaluation, the choice of grain is recorded. For audio resynthesis, the new set of grains is output with a 50ms linear crossfade between grains.

The NN search uses the standard Euclidean distance, facilitated using a k -d tree data structure [28]. Note that the timbre features are normalised for each excerpt, meaning the NN search is in a normalised space rather than the space of the raw feature values.

In both the NN and XAMRT lookup there is an issue of tie-breaking. More than one source grain could be retrieved – at the minimum distance from the query (for NN) or in the leaf node retrieved from the query (for XAMRT) – yet we must choose only one. This is not highly likely for NN search (depending on the numerical precision of the implementation) but will occur in XAMRT when mapping from a small to a large dataset, since the tree can grow only to the size allowed by the smaller dataset. Additional

¹ <http://archive.org/details/xamrtconcat2010>

criteria (e.g. continuity) could be used to break the tie, but for this experiment we keep the design simple and avoid confounding factors by always choosing the grain from the earliest part of the recording in such a case.

4.4 Evaluation method

The ultimate evaluation of musical synthesis techniques is through listening tests; however we defer this to later work, when we plan to incorporate the technique into more complete synthesis systems. For development and comparison purposes it is particularly helpful to have objective measures of success. It is natural to expect that a good concatenative synthesiser will make wide use of the “alphabet” of available sound grains, so as to generate a rich as possible output from the limited alphabet. Here we develop this notion into an information-theoretic evaluation measure.

Communication through finite discrete alphabets has been well studied in information theory [29]. A key information-theoretic quantity is the (Shannon) *entropy*, defined for a discrete random variable X taking values from an alphabet \mathcal{A} as

$$H(X) = - \sum_{i=1}^{|\mathcal{A}|} p_i \log p_i \quad (4)$$

where p_i is the probability that $X = \mathcal{A}_i$ and $|\mathcal{A}|$ is the number of elements in \mathcal{A} . The entropy $H(X)$ is a measure of the information content of X , and has the range

$$0 \leq H(X) \leq \log |\mathcal{A}| \quad (5)$$

with the maximum achieved iff X is uniformly distributed.

If the alphabet size is known then we can define a normalised version of the entropy called the *efficiency*

$$\text{Efficiency}(X) = \frac{H(X)}{\log |\mathcal{A}|} \quad (6)$$

which indicates the information content relative to some optimised alphabet giving a uniform distribution. This can be used for example when X is a quantisation of a continuous variable, indicating the appropriateness of the quantisation scheme to the data distribution.

We can apply such an analysis to our concatenative synthesis, since it fits straightforwardly into this framework: timbral expression is measured using a set of continuous acoustic features, and then “quantised” by selecting one grain from an alphabet to be output. It does not deductively follow that a scheme which produces a higher entropy produces the most pleasing audio results. However, a scheme which produces a low entropy will tend to be one which has an uneven probability distribution over the grains, and therefore is likely to sound relatively impoverished – for example, some grains will tend to be repeated more often than in a high-entropy scheme. Therefore the efficiency measure is useful in combination with the resynthesised audio results for evaluating the grain selection scheme.

Query type	Efficiency (%)
Nearest neighbour	70.8 ± 4.4
XAMRT	84.5 ± 4.8

Table 2. Experimental values for the information-theoretic efficiency of the lookup methods. Means and 95% confidence intervals are given. The improvement is significant at the $p < 0.000001$ level (paired t -test, two-tailed, 19 degrees of freedom, $t = 12.47$).

4.5 Results

We applied the concatenative synthesis of Section 4.3 to each of the 20 pairwise combinations of the 5 audio excerpts (excluding self-to-self combinations, which are always 100% efficient) using each of the two lookup methods (NN and XAMRT). We then measured the information-theoretic efficiency (6) of each run. Table 2 summarises the efficiencies for each lookup method. Our method is seen to be significantly better than the NN search, improving efficiency by over 13 percentage points.

Audio examples of the output are available online.¹ Note that the reconstructed audio examples sound rather unnatural because the experiment is not conducted in a full concatenative synthesis framework. In particular we use a uniform grain duration of 100ms and impose no temporal constraints, whereas a full concatenative synthesis system typically segments sounds using detected onsets and includes temporal constraints for continuity, and therefore is able to synthesise much more natural attack/sustain dynamics [25].

Such factors mean our audio outputs are tricky to judge by listening, and it is not quite clear how far the advantage in efficient use of grains translates into an improved perceptual richness of the output – i.e. into improvements in the timbral analogies made. Nevertheless, our method shows promise as the timbral component of a multi-attribute search which could potentially be used in concatenative synthesis, as well as other applications requiring timbral search from audio examples (e.g. query-by-example [4]).

5. CONCLUSIONS AND FURTHER WORK

We have developed a nonparametric technique able to learn associations from one unlabelled data distribution to another defined on the same space, assuming similarity in structure of the data distributions but accounting for differences in location and shape. This provides a robust and efficient way to map timbre trajectories from one sound source onto timbre trajectories to be performed with a different sound source, making good use of the timbral variation available in the latter. In experiments with a simplified concatenative synthesiser, we have demonstrated that it makes significantly better use of the source material than a nearest-neighbour search.

Future work would integrate this approach into a full concatenative synthesis framework, and supplement the objective tests with listening tests. We also intend to apply the technique to other types of synthesis to control them by audio input.

6. REFERENCES

- [1] D. Schwarz, "Current research in concatenative sound synthesis," in *Proceedings of the International Computer Music Conference (ICMC)*, pp. 9–12, 2005.
- [2] T. Jehan, "Event-synchronous music analysis/synthesis," in *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx-04)*, (Naples, Italy), pp. 361–366, 2004.
- [3] B. Sturm, "Adaptive concatenative sound synthesis and its application to micromontage composition," *Computer Music Journal*, vol. 30, no. 4, pp. 46–66, 2006.
- [4] M. LeSaffre, D. Moelants, M. Leman, B. De Baets, H. De Meyer, G. Martens, and J.-P. Martens, "User behavior in the spontaneous reproduction of musical pieces by vocal query," in *Proceedings of the 5th Triennial ESCOM Conference*, (Hannover, Germany), pp. 208–211, 2003.
- [5] V. Verfaillie, U. Zölzer, and D. Arfib, "Adaptive digital audio effects (A-DAFx): a new class of sound transformations," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 1817–1831, 2006.
- [6] ANSI, *Acoustical Terminology*. No. S1.1-1960, New York: American National Standards Institute, 1960.
- [7] J. M. Grey and J. W. Gordon, "Perceptual effects of spectral modifications on musical timbres," *Journal of the Acoustical Society of America*, vol. 63, no. 5, pp. 1493–1500, 1978.
- [8] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes," *Psychological Research*, vol. 58, no. 3, pp. 177–192, 1995.
- [9] A. Caclin, S. McAdams, B. K. Smith, and S. Winsberg, "Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones," *Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 471–482, 2005.
- [10] J. A. Burgoyne and S. McAdams, "A meta-analysis of timbre perception using nonlinear extensions to CLASCAL," in *Sense of Sounds* (R. Kronland-Martinet, S. Ystad, and K. Jensen, eds.), vol. 4969/2009 of *Lecture Notes in Computer Science*, ch. 12, pp. 181–202, Berlin: Springer, 2009.
- [11] J. M. Grey, "Timbre discrimination in musical patterns," *Journal of the Acoustical Society of America*, vol. 64, no. 2, pp. 467–472, 1978.
- [12] S. Lakatos, "A common perceptual space for harmonic and percussive timbres," *Perception & Psychophysics*, vol. 62, no. 7, pp. 1426–1439, 2000.
- [13] J.-J. Aucouturier and F. Pachet, "Improving timbre similarity: how high's the sky?," *Journal of Negative Results in Speech and Audio Sciences*, vol. 1, no. 1, 2004.
- [14] D. Schwarz, *Data-Driven Concatenative Sound Synthesis*. PhD thesis, IRCAM, Paris, France, Jan 2004.
- [15] D. Deterding, "The formants of monophthong vowels in Standard Southern British English pronunciation," *Journal of the International Phonetic Association*, vol. 27, no. 1, pp. 47–55, 1997.
- [16] E. Chávez, G. Navarro, R. Baeza-Yates, and J. L. Marroquín, "Searching in metric spaces," *ACM Computing Surveys*, vol. 33, pp. 273–321, 2001.
- [17] J. Wouters and M. W. Macon, "A perceptual evaluation of distance measures for concatenative speech synthesis," in *Proceedings of ICSLP'98*, vol. 6, (Sydney), pp. 2747–2750, Nov 1998.
- [18] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Wadsworth Statistics/Probability Series, Wadsworth Inc, 1984.
- [19] F. Questier, R. Put, D. Coomans, B. Walczak, and Y. V. Heyden, "The use of CART and multivariate regression trees for supervised and unsupervised feature selection," *Chemometrics and Intelligent Laboratory Systems*, vol. 76, no. 1, pp. 45–54, 2005.
- [20] G. Lugosi and A. Nobel, "Consistency of data-driven histogram methods for density estimation and classification," *Annals of Statistics*, vol. 24, no. 2, pp. 687–706, 1996.
- [21] C. E. Brodley and P. E. Utgoff, "Multivariate decision trees," *Machine Learning*, vol. 19, no. 1, pp. 45–77, 1995.
- [22] J. Gama, "Functional trees," *Machine Learning*, vol. 55, no. 3, pp. 219–250, 2004.
- [23] S. R. Searle, G. Casella, and C. McCulloch, *Variance Components*. Wiley-Interscience, online ed., 2006.
- [24] S. T. Roweis, "EM algorithms for PCA and SPCA," in *Advances in Neural Information Processing Systems (NIPS)*, (Denver, Colorado), pp. 626–632, 1998.
- [25] E. Maestre, R. Ramírez, S. Kersten, and X. Serra, "Expressive concatenative synthesis by reusing samples from real performance recordings," *Computer Music Journal*, vol. 33, no. 4, pp. 23–42, 2009.
- [26] G. Peeters, "A large set of audio features for sound description," tech. rep., IRCAM, 2004.
- [27] J. McCartney, "Rethinking the computer music language: SuperCollider," *Computer Music Journal*, vol. 26, no. 4, pp. 61–68, 2002.
- [28] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, no. 9, pp. 509–517, 1975.
- [29] C. Arndt, *Information Measures*. Springer, 2001.