PRINCIPAL COMPONENT ANALYSIS OF RASTERISED AUDIO FOR CROSS-SYNTHESIS

Jeremy J. Wells

Audio Lab, Department of Electronics,
University of York, YO10 5DD, UK.
jjw100@ohm.york.ac.uk

Gregory Aldam[*]

Department of Music,
University of York, YO10 5DD, UK.

ABSTRACT

This paper describes a system for cross synthesis of rasterised time-domain audio. Rasterisation of the audio allows alignment of the macroscopic features of audio samples of instrument tones prior to principal component analysis (PCA). Specifically a novel algorithm for straightening and aligning rastogram features has been developed which is based on an interactive process incorporating the Canny detection algorithm and variable resampling. Timbral cross-synthesis is achieved by projecting a given instrument tone onto the principal components derived from a training set of sounds for a different tone. The alignment algorithm improves the efficiency of PCA for resynthesizing tones.

1. INTRODUCTION

Principal component analysis (PCA) is a technique which describes sets of data in terms of the eigenvectors, and corresponding eigenvalues, of its covariance matrix. The first principal component is the eigenvector with the highest eigenvalue, the second principal component (PC) is the eigenvector with the next highest eigenvalue and so on [1]. This process allows identification of features which are common amongst the data. This can be useful for data compression (eigenvectors with low eigenvalues can often be discarded with little detrimental effect on the quality of the representation of the data) and for identification (a specific profile of eigenvalues for a given set of eigenvectors, may give a strong indication that some data 'belongs' to a particular entity e.g. image data may belong to a particular face).

PCA has been applied in many areas from facial recognition [2] to the frequency-domain synthesis of musical instrument tones [3]. Recent work has demonstrated how the rasterisation of one-dimensional audio signals into two-dimensional image-like representations can reveal useful visual analogies with the audio, enabling compact audio display and the application of image processing techniques [4]. This paper details work that draws together these ideas in a system which describes time-domain audio in terms of its PCs and uses this representation as the basis for cross-synthesis between different instrument tones. It differs from previous applications of PCA to audio in that it employs rasterisation and operates in the time, rather than frequency domain. Whilst PCA does not itself require rasterisation (in fact when images are subject to PCA they are de-rasterised), good temporal alignment between different musical instrument sounds is needed for the most compact representations of the data and techniques similar to those applied to image data, such as faces for example, to align features are adopted here for audio.

The paper is organised as follows. A brief summary of PCA and an application of it to image data is given in the next section. Section 3 gives an overview of how PCA is applied to time domain audio in order to produce hybrid instrument sounds via cross-synthesis. The feature alignment which is performed prior to PCA is described and assessed in Section 4. Results for different instrument tones are presented in Section 5 and conclusions given in Section 6.

2. PRINCIPAL COMPONENT ANALYSIS

We consider sets of data which are of the same size and each of which represent a particular entity. For example each set of data may be derived from an image of a human face, or a digital recording of an instrument tone. In many texts on PCA a single data set is referred to as an observation of a set of N variables, where N is the size of the data set. When PCA is used for data compression the idea is to produce fewer, new sets of data which can then be linearly combined to reproduce the each of the original data sets as closely as possible. The best fidelity to the original data is achieved, for a given compression ratio, when the new data sets are as highly correlated with each of the original data sets as possible, but are orthogonal to each other.

The implementation of PCA adopted here is that employed for determining so called 'eigenfaces' from a set of face images in [2]. Here we are interested in determining 'eigensounds' from a set of recordings of a particular type of musical instrument, the 'training set'. Images require derasterisation prior to PCA, for example an N by N pixel image is transferred to an N^2 vector. Usually this would not be a requirement for single-channel audio recordings but the pre-processing prior to PCA described in the next section of this paper requires rasterisation, so in this case derasterisation is also required for audio. Each vector representing a single sound is then centred by subtraction of the mean vector.

[*] Mr Aldam is now with Detica Information Intelligence.

The mean-centred vectors are then combined, each vector forming a column of the N by M matrix, $\mathbf{A}$, where M is the number of data sets (sounds) and N is the number samples in each set. Next the covariance matrix $\mathbf{C}$ is calculated:

$$\mathbf{C} = \mathbf{A}\mathbf{A}^{\mathrm{T}} \tag{1}$$

The eigenvectors of $\mathbf{C}$ are the eigensounds and the contribution of each to the complete set of sounds is given by its corresponding eigenvalue. The first PC of the set is the eigensound with the highest eigenvalue, the next is that with the second highest eigenvalue and so on.

The difficulty with (1) is that the size of $\mathbf{C}$ is N by N which, for instrument sounds, could be several seconds long. At typical audio sample rates, such as 44.1 kHz, finding the eigenvectors and their associated eigenvalues may well be computationally intractable. An alternative approach is to consider the eigenvectors $\mathbf{v}_i$ of the M by M matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ whereby it can be shown that $\mathbf{A}\mathbf{v}_i$ are the eigenvectors of $\mathbf{C}$ [2]. Any sound from the training set can be reconstructed by determining a set of weights to be applied to each eigensound. These weights are determined by projection of the original sound (after mean-centring) onto each eigensound in turn. Reconstruction is achieved by:

$$\mathbf{s}_i = \left(\sum_{m=1}^{M} w_{m,i} \mathbf{u}_m \right) + \overline{\mathbf{s}} \tag{2}$$

where $\mathbf{s}_i$ is the ith sound in the training set, $\mathbf{u}_m$ is the nth eigensound, $w_{m,i}$ is its weight for the ith sound and $\overline{\mathbf{s}}$ is the mean vector.

3. PCA FOR HYBRIDISATION OF TIME-DOMAIN AUDIO DATA

3.1. Audio applications of PCA

Applications to date of PCA to audio have focussed on the spectra of audio, particularly for determining a reduced set of ‘basis spectra’ for multiple wavetable synthesis (MWS) [5]. MWS uses a fixed set of short waveforms to reproduce complex sounds which change dynamically by the alteration of the weights applied to each. The weighting is based on the short-time Fourier transform (STFT) analysis of a target sound, as the target short-time spectrum changes so the weights of the wavetables are selected to best match this. Where memory is at a premium, such as in mobile phones, it is desirable to reduce the number of spectra which are stored in order to represent that of a target sound and how it evolves over time.

Various techniques have been employed to find a suitable reduced set of basis spectra for controlling the amplitudes of the wavetables. In [5] the use of a genetic algorithm (GA) and PCA to select the spectra were compared. Although, using a numerical measure of relative spectral error, they found that the GA was better able to accurately synthesize a recorded sound using fewer wavetables, PCA was faster and offers the potential to be fully automated. Various modifications to their method have been proposed. In [6] complex-valued PCA was used to obtain (i) a set of complex basis spectra, the magnitude of which were used to represent spectral amplitude information and (ii) a set of complex envelopes the magnitude of which were used to represent the overall envelope of the sound and the phase of which introduced deviations from pure harmonicity which, for example, result in the successful representation within the PCs of a flute tremolo. Recent work has improved the quality of real-valued PCA by normalisation of the spectral energy in each frame prior to formation of the covariance matrix [3]. PCA has also been applied to features (e.g. intensity, frequency centroid), for automatic sound classification. For example, in [7] a system for differentiating between silence, speech, music and noise is described.

3.2. PCA of time domain representations of audio

The prime motivation for the work presented in this paper was to develop a novel technique for cross synthesis of instrument timbres. This application of PCA for audio has received some attention but this has been focussed in the spectral domain. For example, in [8] timbral interpolation was achieved by moving through a timbre space whose three axes were defined as the first three PCs derived from a concatenation of several different tone spectra. In [9] a cross-synthesis method which applied the amplitude envelopes of the PCs derived from the spectra of one tone are applied to those of another. These examples indicate that PCA is a promising technique for the creative transformation of audio. However, to date no attention appears to have been paid to how PCA might be applied to producing hybrid timbres from time-domain representations of instrument tones. This is the focus of the work presented here and the following sections of this paper describe and evaluate a method for doing this.

4. FEATURE ALIGNMENT OF TIME-DOMAIN INSTRUMENT TONES FOR PCA

4.1. Rasterisation of audio

Rasterisation is the mapping of data into a two dimensional (2D) space by scanning across the space (usually left to right) in single lines, flying back (right to left) to begin mapping from left to right a line directly below the first and so on. Once the last line of the 2D space has been filled with data, one frame is complete and the fly-back is now from bottom-right to top-left to begin filling the 2D space with a second frame, line by line. Perhaps the most well-known example of this is the physical process performed inside a cathode ray tube, for the representation of moving images on television screens etc.

Rastogram is the term used to describe the raster-scanning of one-dimensional audio into a two-dimensional space. It has been proposed as tool for timbre visualisation, sound analysis and filter design. As a process it is invertible (2D images can be de-rasterised to produce time-domain audio signals) and an intuitive relationship exists between certain textures in images and the sounds produced by sonification of images processed using different visual filters [4]. The rasterisation of audio is defined in [4] as the mapping of one audio sample to one pixel by representing its value as the brightness in greyscale. Whilst rastograms are a time-domain representation, they do offer useful insight into the pitch and harmonic content of sound. If the line width (number of audio samples mapped to each line) is an integer multiple of a sinusoidal component within the sound then vertical stripes will be seen (black for the positive part of the cycle, fading into white for the negative cycle). For many sounds, there will be pitch fluctuations over time and a certain degree of inharmonicity between partials. These aspects can often be clearly seen in rastograms as curved or diagonal lines. What is depicted in such visualisations of audio is very similar to the mapping of analogue audio into the single spiralled groove of a phonographic disk.

4.2. Feature alignment of rasterised audio

PCA does not require rasterisation, in fact de-rasterisation is required before PCA can begin. The purpose of rasterising time domain audio for this application is that it allows similar feature alignment techniques to those used for images prior to PCA. For example, if using PCA for the analysis of faces it usually desirable for common features (e.g. nose, chin and eyes) to be in the same position for each analysed image. If these common features are not in the same position across the set of images then this commonality between images, and therefore 'eigenfaces' is lost, meaning that essential positional information is spread across a larger number of PCs than is necessary.

When using PCA for sound spectra it is usually important that there is alignment in the pitch (fundamental frequency). When considering an instrument tone in the time domain it is important that different portions of the envelope of the sound are aligned and that variations in pitch and amplitude are either aligned or removed prior to analysis. Rasterisation offers the potential for this kind of alignment. The following sub-sections outline the methods employed here.

4.3. Fixed resampling of instrument tones to match period to rastogram width

Inspection of how a rastogram changes as the row (or line) number changes provides visual information about that tone itself varies over time. This visual information is confused if the width of the rastogram is not matched with the time period (the reciprocal of the fundamental frequency) of the instrument tone since vertical alignment is lost between rows. In such cases the rastogram gives the impression that there has been some change in the time domain waveform even where there has been none. The simplest solution to this problem is to adjust the width of rastogram so that the number of pixels in a single line is the same as the number of samples in the period of the instrument tone. This offers an improvement over the unaligned case, but where the time period is not exactly an integer number of samples then there will be some diagonal skewing of features that would be vertically aligned if the period of the tone perfectly matched the width of the rastogram. A straightforward solution to this is to resample the audio. This is achieved via the standard procedure of up-sampling by one integer factor p followed by decimation by a second factor q, combined with low-pass filtering to yield a change in the sampling rate of p/q [10]. Provided $p/q > 1$ this process is reversible since no information is lost as a result of the low-pass filtering. The top panel of Figure 1 shows three rasterised instrument tones where the line width has been set to best approximate the mean period. The bottom panel of the same figure shows rastograms of the same tones which have been resampled in order to produce the best vertical alignment of features. The best ratio p/q was determined by trial and error. Time is indicated on the vertical axes of these plots, however for aesthetic reasons the axes of subsequent plots are not labelled or quantified.

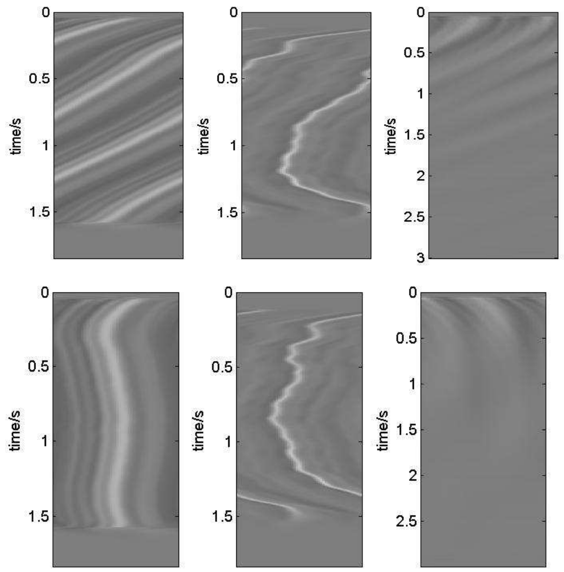

Figure 1: *Rastograms of single tones produced by, from left to right, saxophone, trumpet and electric piano. The top panel shows alignment by matching line width to time period, the bottom panel shows alignment by resampling audio data.*

4.4. Edge detection and gradient estimation for variable resampling of instrument tones

Whilst the method described in the preceding sub-section does improve the visualisation of the instrument tones it assumes pitch stationarity and periodicity. To deal with mild inharmonicity and variations in pitch an algorithm which is able to align according individual dominant components and is able to track variations in such components over time is needed. Since the tones are represented in rastograms this can be achieved by edge detection and straightening. The following procedure is employed after the instrument tone has been resampled at a fixed rate to give the best vertical alignment.

For edge detection the Canny algorithm is used [11]. This can reduce the number of false positives (pixels incorrectly identified as belonging to an edge) and improve the localisation of correctly identified edges (when compared with simpler edge detectors) [11]. The detector begins by smoothing the data with a Gaussian filter to reduce the impact of individual pixels which are particularly noisy. Following this the gradient magnitude and direction for each pixel is calculated from the first derivative of intensity. A process of thinning the probable edge areas then takes place by the suppression of points which are not local maxima in the direction of the gradient. Finally thresholding with hysteresis along with connectivity analysis is applied to detect and link edges.

In order to use this to determine varying resampling ratios to be applied when moving from top to bottom, the rastogram is divided into horizontal strips, each containing *n* lines. Within each strip the Canny algorithm is used to find the longest edge in the strip. The average local gradient *s* of the strip is defined as being the difference (in pixels) between the positions (in the horizontal direction) between the top and bottom of the longest edge. If *s* is positive then the current strip is appended with *s* samples from the next strip, if *s* is negative then *s* samples are transferred from the end of the current strip to the beginning of the next strip. The strip is then resampled by the ratio *v* where:

$$v = \frac{Wn + s}{Wn} \tag{3}$$

where *W* is the width of the rastogram in pixels.

This entire procedure is then repeated until the entire rastogram has been processed. Best results are obtained by iteratively performing this process with differing values of *n*. This resampling and appending process can be reversed once analysis/resynthesis has taken place, provided a resampling 'map' is retained. The top panels of Figure 2 show the vibrato being progressively removed from an oboe tone by four iterations of this algorithm, the bottom panels show the same process being performed on a violin tone. The vertical axis is shown in pixels and it is repeated for the last panel in the figure since this algorithm can lead to a variation in the height of the rastogram. Where $v < 1$ there may be a loss in quality due to the low pass filtering employed in the resampling process. However this loss of quality can be prevented by oversampling the entire rastogram before performing variable resampling. The values of *n* for each successive iteration in the figure (moving from left to right) are 18, 13, 9 and 6 respectively.

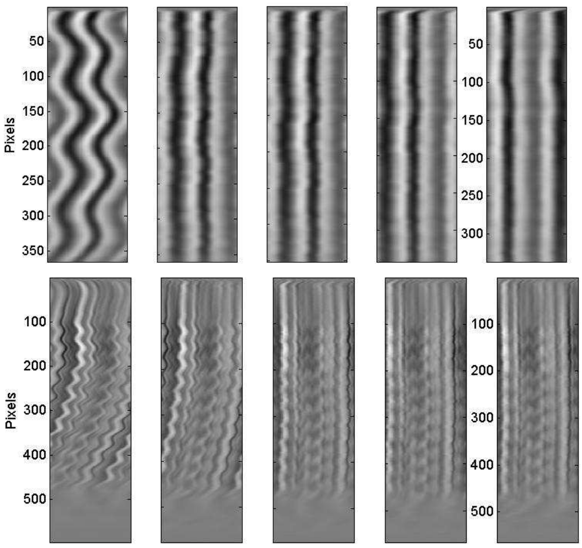

Figure 2: *Rasterised instrument tone, oboe (top) and violin (bottom), variably resampled after edge detection with (from right to left) one, two, three and four iterations.*

Further vertical feature alignment between rastograms is achieved by padding the beginning and end of the tone with zero samples so that the darkest column is centered in the image. This process is illustrated in Figure 3.

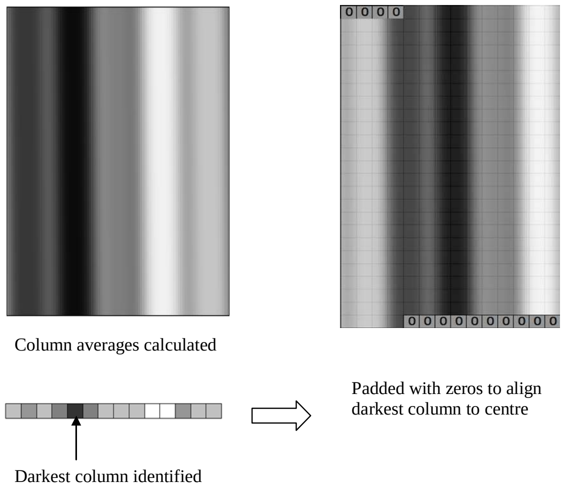

Figure 3: *Column alignment of rastograms*

4.5. Envelope alignment by row insertion or deletion

Having aligned line-to-line features the next stage is to ensure that the rastograms each have the same number of lines. This is achieved by the insertion or removal of entire lines from the steady-state part of the rastogram. This simple approach can be employed since the alignment already performed will ensure that there is high similarity and synchronisation between lines. A basic assumption that the steady state of each tone begins 30% into its duration and is effective for the instrument tones used here, although to make this process more general a better steady-state detection algorithm is needed. To determine the position of the end of the steady state portion the sample values for each line are averaged to derive an amplitude envelope. Once the row amplitude has fallen below one half of the mean amplitude for whole rastogram then the steady-state is considered to have finished. Once the position of the steady-state portion of the tone has been identified it is then expanded, or reduced, to achieve a standard number of rows for the tone. This standard number of rows can be determined by the user and is dependent upon the length of tones being analysed. After this process each rastogram should have the same number of lines.

Figure 4 shows this row insertion/removal (right hand panel) after straightening (middle panel) of a rasterised instrument tone (left panel) for a violin (top) and trumpet (bottom) tone. Provided only insertion is performed and a map of the insertion points is retained this process can be undone, if required.

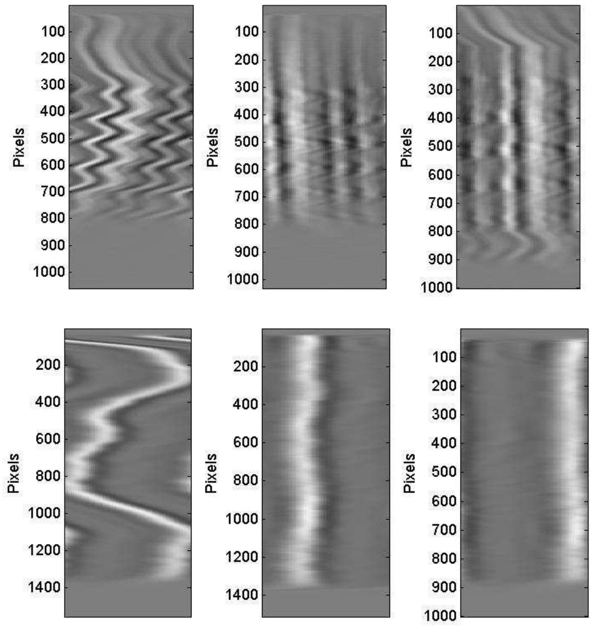

Figure 4: *Rastograms of original instrument tones (left), variably resampled (middle) and steady-state aligned (right), violin tone (first panel) trumpet tone (second panel).*

4.6. Summary of rastogram preparation prior to PCA

A summary of the procedures outlined in this section to align features within, and between, rastograms of different instrument tones is:

1. Trim the audio clip to remove any silence at the beginning and/or end of the signal.
2. Estimate pitch of steady-state portion of signal and use this to determine the width of the rastogram in pixels.
3. Reshape the one-dimensional audio data into a two-dimensional rastogram.
4. Apply iterative edge-detection-based straightening algorithm (described in sub-section 4.4).
5. Perform row insertion/removal to stretch/compress rastogram to desired length.

5. RESULTS

Having produced and aligned instrument tones within rastograms using the methods described in the previous section the audio data is ready for derasterisation and PCA analysis. As stated, the main purpose of this work has been to develop and investigate a novel time-domain cross-synthesis process. This is achieved by projection of the sound of one instrument onto the eigensounds obtained from a set of sounds produced by a different instrument. The purpose of this is to produce a sound that creates a plausible impression of a hybrid of the two instruments. The criteria employed to assess this are:

1. The hybrid sound should be perceived as emanating from a single object.
2. The hybrid should contain some identifiable attributes of both sounds.
3. There should be some sense of acoustic plausibility – an instrument could feasibly exist which might produce such a sound.

This final criterion is, of course, heavily dependent on the differences in the instruments chosen for hybridisation. The purpose of the alignment described in the previous section is to make the PCA more efficient. The following sub-section uses a simple statistic to measure this for two different sounds. Sub-section 5.2 then describes and discusses different cross-syntheses.

5.1. Effect of alignment on PCA efficiency

Figure 5 shows the reconstruction of a horn sound, from different proportions of the 64 eigensounds generated from a training set of which it is a member. It can be seen that, as expected, the fidelity of the reconstruction to the original sound improves as more eigensounds are used. The more efficient a PCA representation of a sound is, the better the reconstruction when using fewer eigensounds. A highly efficient PCA representation of a sound will allow near identical reconstruction from just a few eigensounds. Although the rastograms are very similar, there are some audible differences between the original and the reconstruction from 90% of the eigensounds. Below 70% and the reconstructed sound takes on a more reedy, accordion-like timbre.

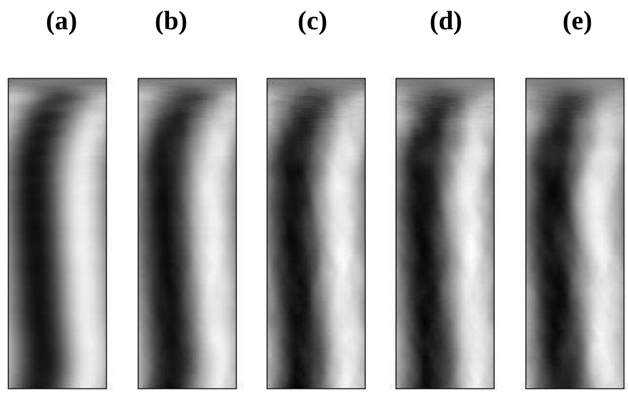

Figure 5: *(a) Original rastogram of a french horn tone. Reconstructions using (b) 90%; (c) 70%; (d) 50%; and (e) 30% of the 64 available eigensounds.*

To assess the effect of the alignment procedures on the efficiency of the subsequent PCA, the mean square of the Euclidean distance between intensity values for each pixel (sample) of an original sound and its reconstruction using eigensounds is presented in Figure 6 for a violin and trumpet. A reduced training set has been used here giving a maximum of 5 eigensounds, which gives perfect reconstruction since the original sound is part of the training set. It can be seen that there is a clear improvement in PCA efficiency for the aligned rastograms over unaligned rastograms. For the violin sound the increase in efficiency, averaged over the 1 to 4 eigensound cases, is 95 %. For the trumpet the average efficiency increase is 48%.

These results show that these alignment processes may well be of use in applications which use PCA of time-domain audio for data compression, although the data required to remove the straightening and inserted rows would need to be stored/transmitted along with the retained eigensounds.

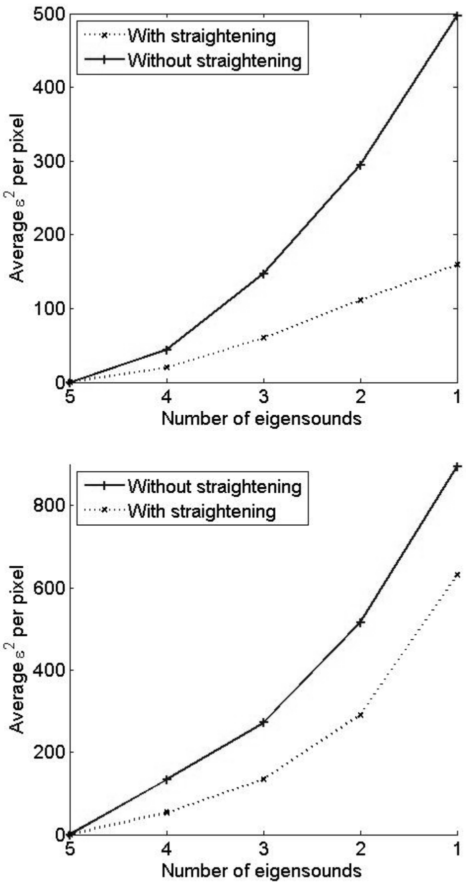

Figure 6: *Square of Euclidean distance between pixel intensities for reconstruction using different numbers of eigensounds, with and without alignment for a violin (top) and a trumpet (bottom).*

5.2. Cross-synthesis

In this sub-section example results, illustrating cross-synthesis are provided. Two training sets of violin sounds have been compiled: one containing 5 tones from different violins, the second containing 20. From these tones cross-synthesized hybrids have been produced by projecting tones from other instruments onto each violin eigensound. The range of pitches in the 5 violin set is C4-E4, the range for the 20 violin set is G4-E4. The instrument sounds used for hybridisation are those of a guitar, trumpet and oboe. Once the hybrid sounds have been constructed from PCA projection onto the violin eigensounds, the structure of the original sound is reintroduced by reversing the alignment that has been carried out prior to PCA, thus returning their original pitch and any frequency modulation characteristics. Audio files for these examples are available online at www.jezwells.org/PCA_audio.

5.2.1. *Guitar*

This is the least successful of the cross-syntheses, due to the vast difference in amplitude envelope between a plucked guitar and a bowed violin. As can be seen from the rastograms in Figure 7a there is very little similarity with the original tone even for the twenty eigensound projection. The result is that the timbre of the violin dominates the hybrid.

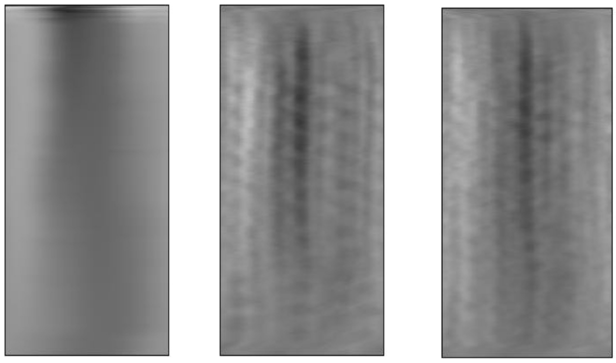

Figure 7a: *Rastograms of original guitar tone (left), reconstruction from eigensounds produced from a 5 violin training set (middle) and a 20 violin set (right).*

5.2.2. *Trumpet*

The trumpet has an amplitude envelope which is closer to that of the violin, particularly in the steady state portion. Whilst the violin timbre does dominate, as the training set size increases from 5 to 20 the spectral envelope approaches that of a trumpet even though the excitation is still very violin-like. The pitch variations in the trumpet tone are retained and the sense of a single acoustic source is strong, making this an interesting and plausible hybrid albeit one which is not perceptually equidistant between the two original instrument types. What is particularly striking about the hybrid produced by this method is the detail which is retained which is often temporally smeared and 'phasey' in spectral hybrids

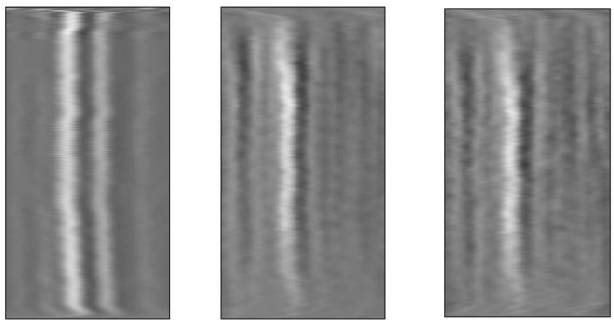

Figure 8b: *Rastograms of original trumpet tone (left), reconstruction from eigensounds produced from a 5 violin training set (middle) and a 20 violin set (right).*

5.2.3. Oboe

A hybrid which is more equally weighted perceptually, particularly for the 20 violin training set, has been produced here however there is an apparent loss of temporal detail which is not the case for the trumpet-violin cross and there is also a chorus-type effect introduced which reduces the sense that there is just a single sounding object. For the 5 violin training set the overall resonant structure is close to that of an oboe but the excitation is violin-like, for the 20 violin training case the excitation sounds more like an equal hybrid of violin and oboe too.

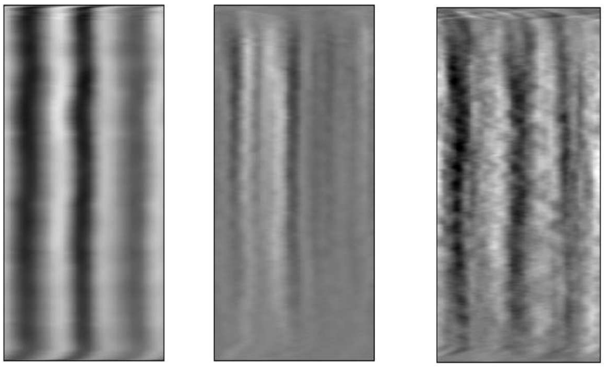

Figure 9c: *Rastograms of original oboe tone (left), reconstruction from eigensounds produced from a 5 violin training set (middle) and a 20 violin set (right).*

6. CONCLUSIONS

We have presented a novel technique for cross-synthesis of instrument tones in the time-domain using PCA. Prior to PCA the audio is rasterised, vertically aligned using edge detection and resampling and horizontally aligned using row insertion/removal during the steady-state portion of the tone envelope. This alignment improves the efficiency of the PCA representation of the training set, meaning that instrument tones in the set are more accurately reconstructed using fewer eigensounds.

The hybrids produced by projecting the sound of an instrument on to the eigensounds produced by PCA analysis of a different instrument demonstrate varying degrees of success. Where the instrument sounds are temporally very dissimilar the training set instrument dominates the hybrid. However where there is greater similarity the cross-synthesis is more successful and plausible, in some cases retaining more temporal detail than tends to be the case with hybrids produced using spectral methods. As the number of eigensounds increases from 5 to 20 there is more scope for the representation of one instrument in terms of the other and a more perceptually equidistant hybrid is usually produced.

There are certainly improvements to the alignment methods used here, particularly achieving horizontal alignment between tones prior to PCA which future work should investigate but the efficacy of this type of alignment for time-domain audio data prior to PCA has been demonstrated and a novel and useful hybridisation process developed.

7. REFERENCES

[1] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, USA, second edition, 2002.

[2] M. Turk and A. Penland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.

[3] S. Wun and A. Horner, "Evaluation of Weighted Principal-Component Analysis Matching for Wavetable Synthesis", *J. Audio Eng. Soc.*, vol.55, no.9, pp. 762-774, 2007.

[4] S. Y. Woon and J. Berger "Application of Raster Scanning Method to Image Sonification, Sound Visualization, Sound Analysis and Synthesis" *Proc. Digital Audio Effects (DAFx'06)*, Montreal, Canada, Sep. 2006, pp. 309-314.

[5] A. Horner et al., "Methods for multiple wavetable synthesis of musical instrument tones", *J. Audio Eng. Soc.*, vol.41, no.5, pp. 336-356, 1993.

[6] T. Rochebois and G. Charbonneau, "Method for multiple wavetable synthesis of musical tones based on complex valued principal component analysis", *Proc. 104th Audio Eng. Soc. Convention*, Amsterdam, May 1998.

[7] N. Nintanda and M. Haseyama, "A new audio signal analyser using PCA and MGD", *Digest of Technical Papers, International Conference on Consumer Electronics*, pp. 1-2, Jan. 2007.

[8] G. Sandell and W Maartens, "Perceptual evaluation of principal-component based synthesis of musical timbres", *J. Audio Eng. Soc.*, vol.43, no.12, pp. 1013-1028, 1995.

[9] T. Rochebois and G. Charbonneau, chapter Music, Gestalt and Computing: Studies in Cognitive and Systematic Musicology (Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence), pp. 375-384, "Cross-synthesis using inverted principal harmonic sub-spaces", Marc Leman Ed., Springer and Verlag, Berlin, Germany, 1997.

[10] J. Watkinson, *The Art of Digital Audio*, Focal Press, Oxford, 2000.

[11] J. Canny, "A computational approach to edge detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.8, no.6, pp. 679-698, 1986.