Guimarães - Portugal

acústica 2004

paper ID:035

RECOGNITION OF NONSTATIONARY SIGNALS WITH PARTICULAR REFERENCE TO THE PIANO

A.C.P. Carrasco[a] , J.K.Hammond[b]

[a,b] *Institute of Sound and Vibration Research ISVR, University of Southampton, UK*

RESUMO: Um dos problemas da transcrição musical automática é o reconhecimento de notas musicais tocadas em simultâneo cujos harmónicos são coincidentes. As técnicas baseadas na transformada de Fourier por definição não conseguem separar este tipo de sinais. Nesta pesquisa estuda-se a possibilidade da utilização de outras transformadas que não sejam baseadas na transformada de Fourier: a Karhunen-Loève Expansion (KLE) e a Singular Value Decomposition (SVD) que demonstrámos serem equivalentes para os sinais não estacionários analisados. Esta pesquisa foi efectuda com sinais musicais de piano modelados e reais.

ABSTRACT: One of the problems in automatic music transcription is the recognition of notes played simultaneously with common harmonics. The Fourier based techniques by definition are not able to separate these kind of signals. In this paper we analyse the possibility of an implementation of a transform that it is not Fourier based. The chosen transforms were the Karhunen-Loève Expansion (KLE) and the Singular Value Decomposition (SVD) that we showed to be equivalent for nonstationary signals. The research is performed over modelled and real piano notes.

1. PIANO NOTES

The recognition of nonstationary signals was specially performed over piano signals that were considered as nonstationary signals. First we built a simple model of the piano notes to study the application of KLE and SVD and then we moved to real piano notes.

From the analysis of piano notes we decided that we could considered them as nonstationary signals and modelled them as transient signals with one frequency component (later we introduced more components), generated according to the equation,

$$x_i(t) = A_i e^{-\alpha_i t} \sin(2\pi f_i t + \phi_i) \qquad (1)$$

where $x_i(t)$ is a possible event, A_i is the amplitude, α_i the damping, f_i the frequency, ϕ_i the phase and t the time that will vary randomly from realisation to realisation. We assign the same probability density function to all the parameters. Each of these parameters A_i, α_i, f_i and ϕ_i is considered to be drawn from independent uniform probability distributions.

2. THE KARHUNEN-LOÈVE EXPANSION (KLE)

The KLE is a mathematical technique used in the treatment of stochastic processes. The KLE is an orthonormal expansion and represents any stochastic signal as a linear combination of uncorrelated functions of a basis set. The signal itself selects the optimal new basis set which is a particular feature of this expansion. The analysis of the KLE theorem shows that it can be applied to any nonstationary stochastic signal.

The discrete version of KLE says that if $X_\mu(t_k)$ is the ensemble of random discrete functions where t_k refers to discrete time and μ is a possible realisation of the signal then:

$$X_\mu(t_k) = \lim_{N\to\infty} \sum_{n=1}^{N} \sqrt{\lambda_n} c_n(\mu)(\Phi_n)_\mu \qquad (2)$$

where $\{\Phi_n\}$ are a set of orthonormal eigen vectors of the covariance matrix R whose elements are given by $(R)_{ij} = E[X_i X_j]$ satisfying: $\sum_{i=1}^{\infty}(\Phi_m)_i(\Phi_n)_i = \delta_{mn}$ (3)

The coefficients of the expansion $c_n(\mu)$ are given by:

$$c_n(\mu) = \left(\sqrt{\lambda_n}\right)^{-1} \sum_{i=1}^{\infty} (\Phi_n)_\mu X_i(t_k) \qquad (4)$$

The eigen vectors of the expansion are the set of the orthonormal functions of the expansion. The correlation matrix is a Hermitian matrix satisfying: $R\Phi_n = \lambda_n \Phi_n$ (5).

In practice the KLE is applied to stationary and weakly stationary random signals in the same way. The procedures for KLE applied to those signals are: estimation of the autocorrelation function (the values are taken from the zero lag to the chosen maximum lag), creation of a square matrix of Toeplitz form from estimated autocorrelation, the eigen decomposition of the autocorrelation Toeplitz matrix. The eigen vectors of the matrix are the vectors of the basis of the KLE and the eigen values are the corresponding energy associated with each vector.

3. APPLICATION OF KLE TO NONSTATIONARY SIGNALS

To apply the KLE to any signal implies the estimation of the autocorrelation function. For nonstationary signals we have to estimate the nonstationary autocorrelation function. For our modelled signal it was possible to determine the theoretical autocorrelation function and compare it with the estimated autocorrelation function by ensemble averaging. The estimation is given by $R = X^T X$ (6), where X is a set of possible realizations of the discrete signal (X is a row matrix). We verified that both autocorrelations were quite similar and could replace the theoretical by the ensemble averaging.

We built the matrix with a set of 100 possible events with 300 samples each (underdetermined system) of a signal using eq.1 which amplitude always was unity, the damping varied in the interval $[-0.3,-0.1]$ and the frequency in the interval $[9.5,10.5]$ Hz and estimated the nonstationary autocorrelation function by ensemble averaging. The absolute values of the nonstationary autocorrelation function of the nonstationary signal (fig.1) show clearly that the matrix has not the Toeplitz structure, as in the stationary case, but is still in the conditions of KLE.

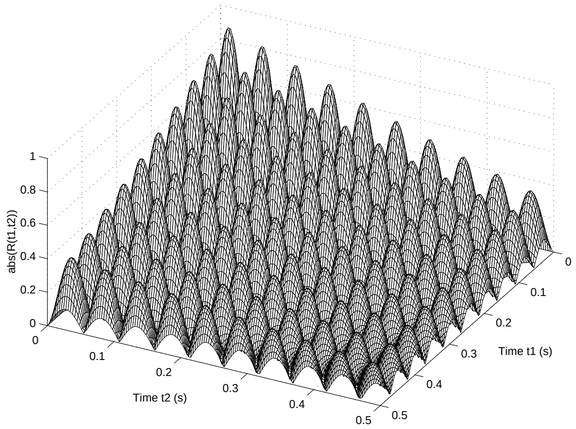

Fig.1 - *Absolute values of the ensemble averaged autocorrelation matrix.*

The eigen vectors of this matrix are the KLE basis set of our the nonstationary signal (fig.2 right). The eigen values of the simulated signal (fig.2 left) are related to the energy of each vector.

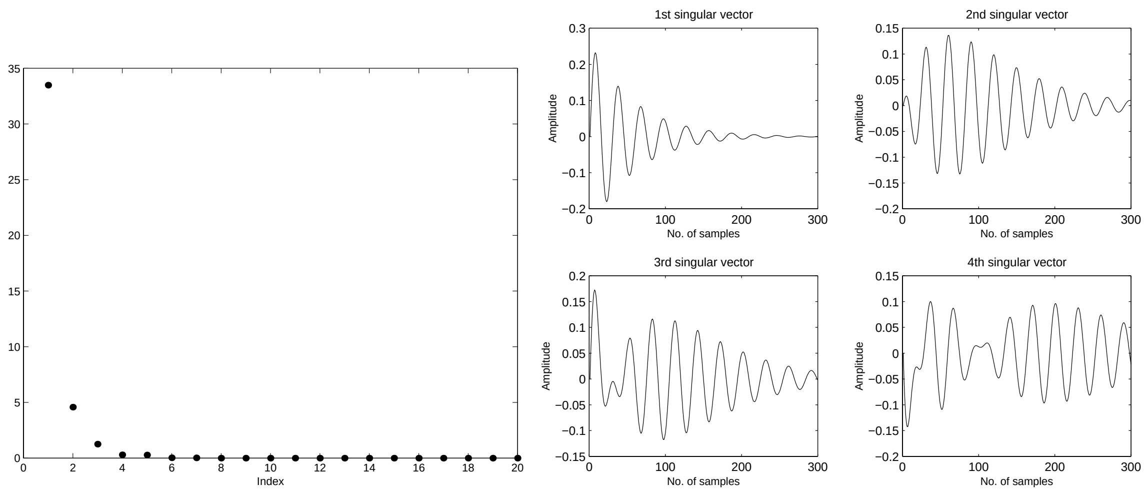

Fig.2 - *Eigen values of the simulated signal for the nonstationary case (left). First four eigen vectors for the simulated signal in the nonsntationary case (right).*

The first singular vector (fig.2 right) has a shape very similar to the original signal. All the transient characteristics of the original signal are present in this vector. The other vectors present modulation and the 3rd and 4th vectors appear to be a version of the 2nd vector moved to the right.

4. THE SINGULAR VALUE DECOMPOSITION (SVD)

The SVD is a mathematical method similar to the KLE also used in matrix decomposition. Although similar, SVD operates over different statistical moments of the signal in the sense that for KLE we need to calculate the autocorrelation function and for SVD we simply need to form the time history matrix. The SVD becomes very convenient for nonstationary data characterised by an ensemble since we do not need the second order statistics as for the KLE.

Any matrix $A(m\times n)$ can be represented in the form of a product of three matrices defined as,

$$A=U\Sigma V^T \tag{7}$$

where $U(m\times n)$ and $V(m\times n)$ are orthonormal matrices. The matrix V^H is a Hermitian matrix (for real matrices, V^H is replaced by the transpose V^T). If the rank of A is $k=\min(m,n)$, the diagonal matrix $\Sigma(m\times n)$ has k nonnegative diagonal elements arranged in descending order,

$$\Sigma = diag\left(\sigma_1,\sigma_2,\ldots,\sigma_k\right) \tag{8}$$

with real $\sigma_i \geq 0$, $i=1,2,...,k$. These elements of the diagonal matrix are the singular values of the matrix A. The columns of matrix U and V are defined as the left and right singular vectors of matrix A.

5. THE RELATION BETWEEN KLE AND SVD

The singular values and the eigenvectors of a matrix can be related. For a matrix $A(m\times n)$ the nonnegative square roots of the eigen values of the matrix product $A^H A$, if $m\geq n$, and of AA^H, if $m\leq n$, are said to be singular values of A (Lütkepohl,1996). (For real data A^H is replaced by the transpose A^T.)

If a matrix A, represented by equation 7, is multiplied by its transpose A^T (left side and right side), the products will be:

$$A^T A=\left(U\Sigma V^T\right)\left(U\Sigma V^T\right)=V\Sigma^2 V^T \quad (9) \quad \text{and,} \quad AA^T=\left(U\Sigma V^T\right)\left(U\Sigma V^T\right)^T=U\Sigma^2 U^T \quad (10)$$

since the matrices U and V are orthonormal and Σ a diagonal matrix.
The ensemble average estimator for autocorrelation of X is computed as:

$$\hat{R}_x\left(t_1,t_2\right)=\frac{1}{n}\sum_{i}^{n}x_i\left(t_1\right)x_i\left(t_2\right) \tag{11}$$

This estimator for the autocorrelation of a matrix X, representing an ensemble of possible realizations of a discrete signal, is equivalent to the product matrix (depending if X is a row or a column matrix):

$$R = X^T X \qquad (6) \qquad \text{or} \qquad R = XX^T \qquad (12)$$

The products AA^T and $A^T A$ represent the autocorrelation matrices of the matrix $A\left(m\times n\right)$ (as mentioned before, depending if A is a column or row matrix. The matrix Σ^2 is diagonal whose elements are the eigen values of the products and correspond to the square of the elements of Σ. If λ_i are the eigen values of the matrix products and λ_i and the singular values of A are ς_i then: $\lambda_i = \xi_i^2 \quad i = 1, 2, 3, \ldots$ (13).

The eq.9 and eq.10 show that KLE and SVD are equivalent for any matrix $A(n\times m)$. The SVD is applied over $A(n\times m)$ while KLE is applied over the covariance matrices AA^T or $A^T A$. Both decompositions have similar vectors and the singular values are the positive square roots of the eigen values.

6. THE KLE AND SVD APPLIED TO SIMULATED NONSTATIONARY SIGNALS

The SVD for nonstationary processes can be applied directly over the time history matrix containing the possible realisations of the signal. We tested in practice the KLE and SVD for signals expressed by eq.1 and concluded these transforms are similar. The singular and eigen values represented are related according to eq. 13. The singular and eigen vectors have similar shapes as was expected according to eqs.9 and 10, although are not exactly the same in some cases. Some singular vectors have symmetric values in relation to the same order eigen vector. The similarities between vectors continue until the singular values (and eigen vectors) become very small. At this point there are no more similarities but the energy associated is very small.

We can conclude that SVD is an equivalent technique to KLE as we saw theoretically and the differences pointed are not significant for our research. The application of the SVD is much simpler, since we will be working directly with the time history matrix without the explicit need to create second order statistics.

7. RECOGNITION PROCESS FOR MODELLED PIANO NOTES

This recognition process for simulated piano notes that we developed and want to implement has the following steps: simulate of a set of time histories with a number of possible events of a signal defined by the frequency representing the pitch, decompose the time history matrix of the signal using SVD, project the signal for recognition in the new spaces defined by the basis set determined by the SVD and identification of the signals.

All the time history matrices have the same dimension as before (100X300). The frequency is the principal variable that defines a simulated signal associated, in real piano notes, with a pitch. A signal is similar to another if they have similar fundamental frequency falling inside a certain interval of error. The resulting singular vectors of the decomposition will be the "reference" of the signal. They are the basis set of a new space where we will project other signals that we want to recognise. We search for similarities between the projected signal and the vectors of the basis set to recognise the notes present in the signal.

On the next figs.3 and 4 we have the examples of the signals composed by five frequency components (harmonics) s_1, s_2 and s_{12} (this signal is composed by the sum of s_1 and s_2 and simulates an octave). They were projected on the basis set of signal s_1 and s_2 to verify if it was possible to detect those notes on the signals and recognise the notes on an octave.

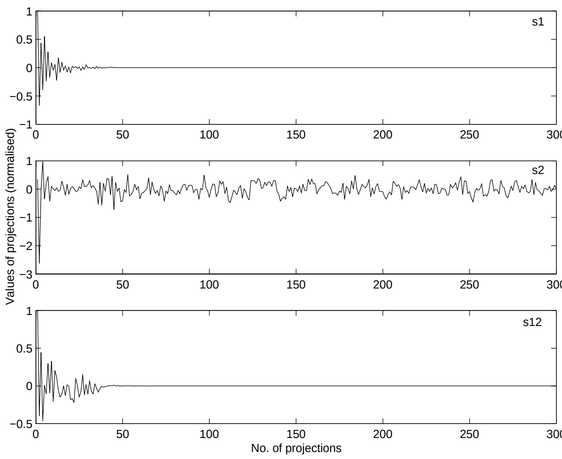

Fig.3 - *Projections of signals* s_1, s_2, *s12 on the basis set* Bs_1.

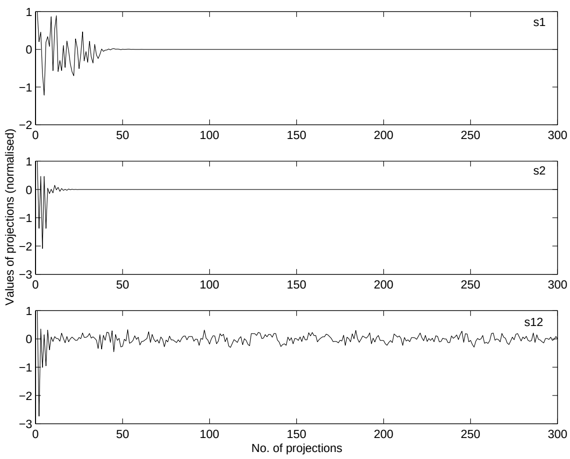

Fig.4 - *Projections of signals* s_1, s_2, *s12 on the basis set* Bs_2.

The norm values for those projections can give more information about those signals. Through the analysis of the results projections itself can give some information about the presence of the note on the signal when there is a maximum value on the first projection (on the most energetic singular vector). That does not always happen and we need to determine the norm of those projections to have more information.

The norm of a signal when projected on its own basis set reaches the total value before any other signal. A very important result was that the identification of an octave was possible using this method as we can see on fig.5.

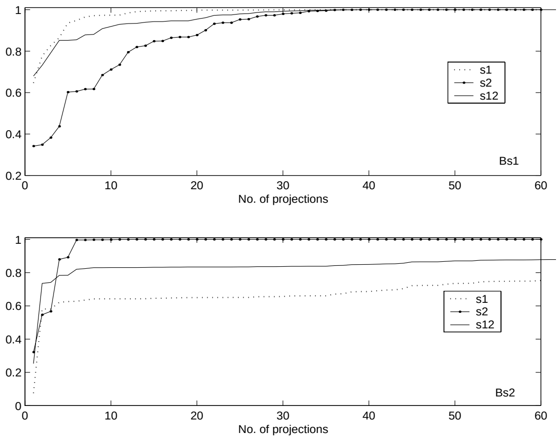

Fig.5 - *Norm of the projections of signal* s_1, s_2 *and* s_{12} *on the basis sets* Bs_1 *and* Bs_2.

8. RECOGNITION OF REAL PIANO NOTES

The recognition process for a real piano, based on the process for simulated notes, notes has the following steps: recording of the possible events of each piano note, determination of the singular vectors of each note, projection of the piano signals on the basis sets defined by the singular vectors of each note, detection of the presence of a piano notes in the musical signal.

For each piano note we recorded a set of 60 events played with different dynamics (and timbre) from *piano* (*p*) to *fortissimo* (*ff*), and with both pedals, *una corda* and *sostenuto* timbre. The number of samples taken from each event took in consideration the number of periods for the recorded notes and the computation time. We took firstly 2000 and then 3000 samples (2000 samples correspond approximately to 0.045s for $f_s = 44100$ Hz, and is enough for the human ear to perceive the pitch of a note played (Gelfand, 1981)). The SVD was applied first to signals aligned at the initial part of the attack and then with delays.

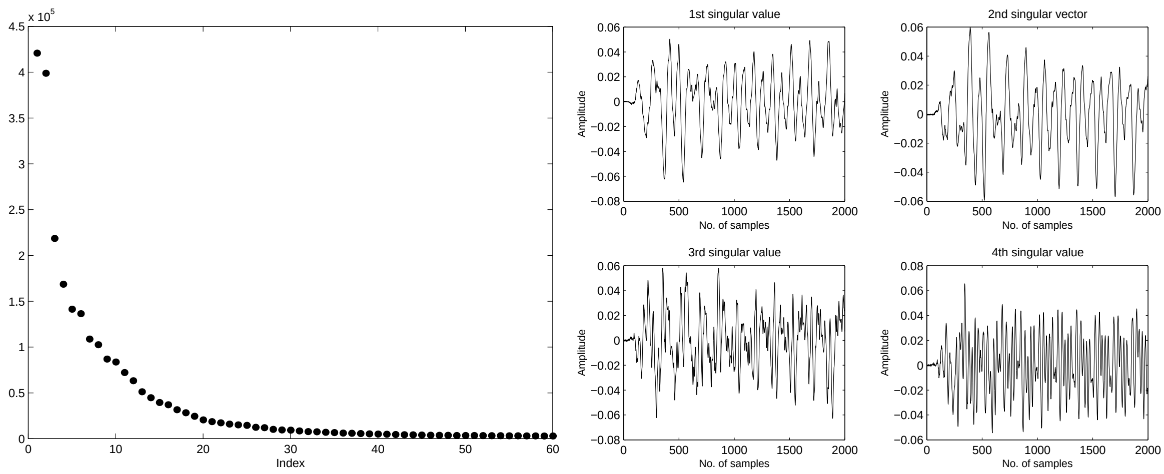

Fig 7- *Singular values (left) and singular vectors (right) of a piano note C262Hz with delays between events.*

The SVD of aligned events shows that the variability of the system is very small and can be reduced to only one dimension. All the dynamics and the use of the two pedals do not give a great variability to the system. Relating this with the results of the model we can conclude that when a piano note is played in several possible ways the shape of the time history continues to be very similar and the most important change is in the amplitude. One of the important variables that can be introduced in the time history matrix is the delay between events. With the inclusion of the delays the dimension of the system increases and the shape of vectors change. This will spread the energy over more singular vectors (fig.7). The exact starting point of a note for recognition and the exact aligning with the singular vectors for projection is not a concern. It is enough to have a value for the starting point that can have some error that fits inside of the delay interval of the events on the time history matrix. It is more convenient to have more significant vectors for the recognition process.

The recognition of all the notes recorded was tested and we saw that the norm of the notes when projected on their own basis set reach first higher values near one. We tested other notes and we conclude that the recognition of single piano notes is fully successful.

It was possible to recognise the notes in an octave and to separate the octave from the note with the lowest fundamental frequency. This recognition depends on the size of the sample and the frequency of the notes involved and because of that we incremented the number of samples from 2000 to 3000. The rate success of this recognition was around 65 % for lower frequencies and 70 % for higher frequencies. For signals composed of two notes separated by a tone the method it is not always able to detect the two notes. From the studied cases the two notes have the maximum values but we cannot infer the number of notes present.

For chords composed by four notes this task becomes much more complex and the results were not enough conclusive. In some cases we detect a phantom note, absent from the signal that can be considered as a common subharmonic of all the notes in a chord (form major chords).

9. CONCLUSION

We showed that SVD could replace KLE and proved to be a goof technique for detecting single notes with a 100% of matching. It is also possible to separate notes with coincident harmonics as the octaves task that is not possible to do with Fourier based techniques. For chords the results are more ambiguous since we detect the notes in the chord and notes that are absent but are subharmonics (they belong to the same harmonic series).

Bibliography

1. Gelfand,S.A. (1981)."Hearing An Introduction to psychological and physiological acoustics", Marcel Dekker, NewYork (1981).
2. Gray,R.M.(1998). "Toeplitz and Circulant Matrices: A review", http://www-isl.stanford.edu/~gray/toeplitz.pdf
3. Hammond,J."Applied digital signal processing," Vol.1 Notes
4. Lutkepohl,H.(1996). "Handbook of Matrices", Wiley& Sons, England 1996
5. Newman,A.J.(1996). "Model Reduction via the Karhunen-Loeve Expansion Part I:An Exposition" http://www.isr.ud.ed