THE USE OF SPECTROGRAPHIC TEMPLATE MATCHING TO IDENTIFY AND CLASSIFY SALIENT SOUND EVENTS IN TENNIS MATCHES

Krzysztof Zienowicz Faculty of C.I.S.M, Kingston University, KT1 2EE, U.K.
Gordon Hunter Faculty of C.I.S.M, Kingston University, KT1 2EE, U.K.
Ahmed Shihab Faculty of C.I.S.M, Kingston University, KT1 2EE, U.K.

1 INTRODUCTION

Some significant events in sports matches occur over the course of a few milliseconds[1] - too quickly to be detected by slow-motion viewing of conventional video captured at 25 or 30 frames-per-second. An audio signal, normally recorded simultaneously with the video, but sampled at a much higher temporal rate, provides a way to detect such events which happen over a very short timescale. However, there are several complications such as background noise, echoes and acoustic latencies (delays due to the speed of sound being several orders of magnitude lower than the speed of light). One of the few examples of the audio signal being employed in the analysis of events in sports matches is the *Snickometer*[2], used in televised coverage of cricket matches to give the commentators and spectators an indication of whether the ball has made contact with the bat, the batsman's glove, arm, (leg-guard) pad or other item of protective equipment, or a combination of these in turn. This can affect the correctness of important decisions by umpires regarding whether or not the batsman should be given "out" in situations where the fielding team claim he should be "out – caught" (where the ball should have hit the bat of glove) or "out – Leg Before Wicket" (LBW), where the ball must have hit the batsman's foot, leg or pad without first having hit the bat. However, at present, the *Snickometer* only examines the "shape" of the sound signal (detected by a microphone in one of the stumps, very close to where the event took place) – looking for profiles characteristic of slight or major bat on ball contacts, or "dull thuds" : damped sounds characteristic of the ball hitting an item of protective clothing – it does not perform any frequency analysis of the sounds.

In this paper, we describe how, by detailed study of both audio and video signals from recordings of championship tennis matches, we have found evidence which suggests that use of "template matching" based on the frequency spectrogram and power of the audio signal – a technique which has been used in the automatic recognition of speech sounds for some time - could lead to the reliable detection and classification of salient sound events - such as racket or bat on ball impacts, the ball clipping the net and bounces of the ball. This could be of benefit to spectators, match officials and coaches both in tennis and other sports - including cricket, baseball and golf. It could also find applications in improving coaching aids and making video games (such as the Nintendo *Wii*) more realistic**.** We are also considering applying our methodology to a wider range of situations, such as monitoring the well-being of sick, infirm or elderly people and to possible applications in security and surveillance. However, such situations are much less controlled environments than championship tennis matches, and applying our procedure to such cases is likely to present new problems and challenges.

2 SALIENT AND AMBIENT SOUNDS IN TENNIS MATCHES

In tennis, there are various "sound events" which occur which are directly related to the play of the game, and various others which get in the way of detecting, analysing and classifying those "game-related sound events". The former category – the salient sounds - includes the sounds produced by the various tennis strokes (including "good" strikes of the ball and mis-hits), by the ball bouncing on the surface of the court, by the ball hitting or "clipping" the net. Many of these are produced by "impulsive" impacts : in a tennis stroke[1], the racket is typically in contact with the ball for

approximately just 5 ms. The other category – the ambient sounds - includes "on-court noise" sounds produced during the game in some peripheral way : echoes, sounds of footsteps, of the ball hitting a back or side wall, "grunts" and similar vocalisations from the players whilst playing (or attempting to play) strokes, applause & "cheering"-type sounds from the audience and speech from the players, match officials, T.V. (or radio) commentators and/or spectators and "bleeps" from the *Cyclops** system used (up to 2006) to determine whether the ball was "in" or "out". The "ambient sounds" may also include background noise from a wide range of sources, such as traffic on nearby roads, sirens of emergency vehicles, road or building work, and/or aeroplanes flying overhead. An audio system can be "calibrated" to allow for the more consistent of these (e.g. general traffic noise) by monitoring the audio signal from on-court microphones while no tennis match is in progress. However, the properties of the more sporadic noises, such as sirens, would have to be specially studied and identified separately. Many of these background noise sounds can be analysed in much the same ways as the on-court noise.

The fact that the speed of sound in air (approximately 340 m/s at sea-level at 20^{o}C) is several order of magnitude lower than the speed of light (about 3 x 10^{8} m/s) results in "acoustic latencies". A standard tennis court[1] is 78 feet (23.774 metres) long and 36 feet (11.077 metres) wide, with diagonals measuring 85.91 feet (26.433 metres). A consequence of this is a microphone placed directly behind a player at one end would receive a sound made by that player approximately 78 ms earlier than it would a sound made at precisely the same time in the opposite corner. This is approximately the duration of 2 frames of conventional video, so care must be taken not to take the time when the sound signal reaches the microphone as the exact time when the actual event producing the sound took place.

Echoes, which are quite prominent in many examples in our dataset, are also consequences of the finite speed of sound. They are due to acoustic reflections from nearby solid objects. Reflections off from solid plane walls, which are often behind the player at one end, will be "specular" and hence lead to strong echoes, but those off irregularly-shaped surfaces (e.g. a crowd of spectators) will tend to be "diffuse" and give weaker echoes, as will those from more distant plane surfaces. It is also possible that certain court surfaces (notably the "hard courts" used in the U.S, and Australian Open Championships) may result in significant echoes, but this factor has not yet been investigated in our study, and the Wimbledon championships are played on grass courts, of which the surfaces are likely to be relatively absorbent acoustically.

In the remainder of this paper, we propose and test a methodology for detecting and distinguishing between these different types of sounds. Initially, we do this by visual inspection of spectrograms and signal power-time profiles. We then go on to show how this process can be automated, with considerable success, for the set of data for tennis matches which we have available.

3 DATASET AND TERMINOLOGY

Our dataset consists of four videos, including the soundtrack, recorded from TV broadcasts covering the Wimbledon lawn tennis championships of 2005, giving over 95 minutes of material in total. We "*ground-truthed*" the footage by marking the occurrence of all impulsive sounds. This gave us over 1400 sound events (including from a total of around 800 tennis strokes) in 14 classes (9 different types of tennis strokes, 5 types of "other sounds") - see Table 1. We also added information about background noise (from the audience, bleep, commentator's speech) and which end of the court the sound was generated ("North" or "South"). The sound was sampled at 48 kHz, which allowed spectrographic analysis in the range 0 – 24 kHz.

*In many major tennis tournaments between around 1980 and 2007, an automated system based on whether or not the ball had interrupted a beam of light - sometimes referred to as *Cyclops* – was used to detect when the ball landed just out of court, or if a service was marginally long or wide. This decision was signalled to the match officials and spectators by a "bleep" noise. However, this system was replaced in the "Grand Slam" tournaments, including Wimbledon in 2007, by the Hawk-Eye[15] system which uses 6 or more special video cameras capable of taking over 500 frames a second.

Code	Meaning	Number in Dataset
S1	First serve	150
S2	Second serve	64
FD	Forehand drive	257
BD	Backhand drive (one handed)	146
BD2H	Backhand drive (two handed)	110
SM	Smash	5
VO	Volley	3
SS	Stop shot	8
LO	Lob	7
EC	Echo	298
BC	Bounce	216
SL	Silence	50
AP	Applause	64
SP	Speech	126

Table 1 : The 14 distinct "sound events" (9 types of tennis strokes, 5 other sounds) which we investigate here, and the number of times each of them occurs in our dataset

The power in an acoustic signal is proportional to the square of the signal intensity. In practice, the average power, or *short-term energy (STE)*, is calculated by integrating the square of the intensity over a short time window. A *spectrogram* is a means of indicating how, at any specified instant of time, the intensity or power in a sound signal is distributed by frequency, and how that distribution varies over time. Spectrograms are produced using a *Fast Fourier Transform (FFT)* or *Mel Cepstral (MFCC)* analysis of the signal over a relatively short time interval[3,4]. The MFCC approach is discussed in more detail in section 5.2 below.

4 PREVIOUS WORK ON THE AUTOMATED DETECTION OF SALIENT SOUNDS

4.1 Work by Other Authors

Previous work on the automated identification of "salient sound events" have followed a variety of approaches, with varying degrees of success. Although some of these have attempted methods based on template matching, these have not always achieved good results. For example, Zhang & Ellis[5] found that an approach based on template matching did not work well in the context of detecting the sound of "dribbling" in basketball. This led to other authors, including Dahyot et al[6], rejecting template matching as a possible method for detecting the sounds of tennis strokes, despite the acoustic environments of tennis and baseball matches being very different – in tennis, many of the salient sound events are of short duration (of the order of 10-20 ms) but separated by intervals of order 1 second, whereas in basketball the ball bounces can occur several times a second. Furthermore, audiences in basketball matches tend to be quite noisy for much of the time. In major tennis championships, the spectators are normally quiet most of the time – except in situations of highly partisan support (e.g. "Henman Hill" at Wimbledon in recent years) ! Although some authors (e.g. Hsu[7] studying golf strokes and Rui et al[8] detecting bat on ball impacts in baseball) have used an approach which has some aspects in common with our "template matching" approach, they have only used the signal power in a very small number of frequency bands[8] or a very small number of MFCC coefficients in conjunction with a neural network[7]. Some authors have only tried to *detect* impulsive sounds[6,9], not classify them. Several studies have used peaks in Short Term Energy (STE) - the signal power averaged over a very short time window[9,10].

Amongst the relatively small number of studies which attempt to classify different types of sounds[11,12,13,14] (rather than just detect salient impulsive sounds), some only attempt to distinguish between radically different sounds[12] (periodic, impulsive, close to monotone or of very limited spectral range) and/or sounds of a small number of distinct classes (6 classes in the case of Dufaux et al[11]). Zhang & Kuo[13,14] used a "hierarchical" approach, using three levels of "coarse",

"intermediate" and "fine" discrimination between sound types. At the "coarse" level, they used the power, zero crossing rate and fundamental frequency of the signal to distinguish between speech, music, environmental sounds and silence. The second level employed Hidden Markov Models (HMMs) to sub-classify each category (e.g. for speech, whether the speaker is an adult male, adult female or a child; for music which genre – classical, jazz, rock, etc. - the sample belongs to). The third stage comprised a querying/retrieval approach to identify the most similar "previously heard" sounds to the current example.

4.2 Our Previous Work

In our previous work on this topic[16], we studied the dependence of the time interval between successive strokes and the acoustic energy from the first of those two strokes. It was found that, as predicted by Newtonian dynamics (assuming that the acoustic energy was proportional to the kinetic energy imparted to the ball in the stroke which produced that sound), the logarithm of the energy was, to a good approximation, linearly dependent on the logarithm of the time until the next stroke. This implies that the energy in the acoustic signal can be a reasonable predictor of the time when the next stroke occurs. Plots of Short Term Energy against Time to next stroke, for the various distinct classes of tennis strokes, are shown in Figure 1.

In the same paper[16], we also considered the statistical distributions of both the acoustic energy and the time interval before the next stroke for various different classes of stroke. Not surprisingly, some strokes (such as first serves and forehand drives) were found to be consistently of high energy and a short time until next stroke, whilst others (such as lobs) were of lower, but more widely varying, energy and longer and widely varying time intervals to the next stroke. This information is summarised in Table 2.

Stroke Type	Mean of T	Mean of STE	SD of T	SD of STE
S1	0.703	2836.3	0.074	977.0
S2	0.966	1363.6	0.158	704.2
FD	1.271	2010.4	0.254	917.7
BD	1.579	1721.3	0.366	1023.6
BD2H	1.275	1849.8	0.166	495.2
LO	2.178	1101.5	0.668	501.0

Table 2 : means and standard deviations of the time before next stroke (T, in seconds) and short-term energy (STE, in arbitrary units determined by the signal amplitude) for various types of tennis strokes in our dataset. See Table 1 for stroke codes

In the present paper, we extend this work by investigating the feasibility of both detecting and correctly classifying tennis strokes and other "salient sounds" automatically, solely on the basis of evidence present in the acoustic signal.

5 METHODS FOR THE ANALYSIS OF "SALIENT SOUNDS" IN TENNIS

5.1 Spectrographic Profiles

As an initial approach, in order to get a better appreciation of the problem we were likely to face, we studied (by visual inspection) the spectrograms of the various types of "salient sounds" we wished to be able to detect and classify. Examples of these are given in Figure 2 below. At least to the human eye, we found a large degree of consistency between the spectrograms of strokes of the same type, and other sounds (such as applause and speech) also had characteristic features. For example, speech (Figure 2(m)) shows distinctive bands (indicating the "formants" of vowels and other voiced sounds) lasting of the order of hundreds of milliseconds, whereas applause (Figure 2(l)) from a sizeable audience shows a "fuzzy" spectrogram with the signal power spread over a

wide range of frequencies (approximately 0-12 kHz, with most energy between 0 and 3kHz) but may last for several seconds. However, it was found that many strokes (e.g. serves, forehand drives, both one and two handed backhand drives – Figure 2(a) to (e)) had spectrograms which appeared rather similar to each other. It was thus not easy to classify what type of stroke had been played from inspection of its spectrogram alone.

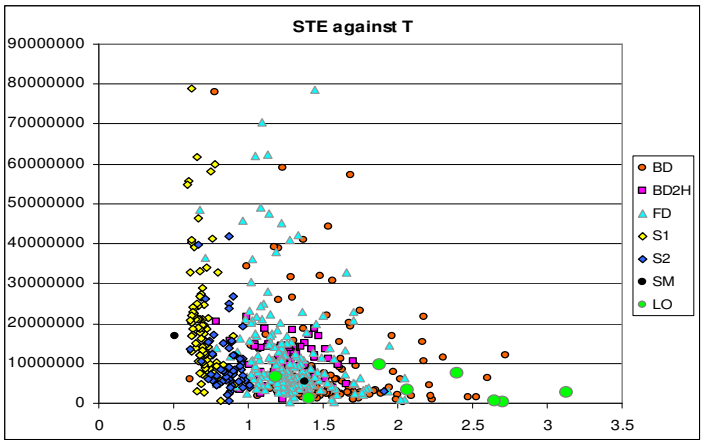

Figure 1 : Graph showing the inter-relationship between the acoustic Short Term Energy (STE – vertical axis) of a tennis stroke and the time interval (T – horizontal axis) in seconds before the next stroke., for various different classes of strokes. The codes for the different types of strokes are as specified in Table 1.

The example (Figure 2 (k)) where the ball has clipped the net shows some interesting features : the "bleep" is a very simple periodic sound consisting of a fundamental at about F = 3 kHz with a couple of weaker harmonics at 2F and 3F – indicated by the horizontal lines across the spectrogram at these frequencies. Furthermore, the spectrogram shows that the microphone used to record the acoustic signal must be nearer to the loudspeaker for the Cyclops bleep than to the point where the ball clipped the net – the "bleep" shows up on the spectrogram a short, but detectable, time before the impulsive sound of the ball on net contact ! Several prominent echoes are also visible in that example. The example in Figure 2 (h) was of a weak lob played during a large amount of noise (including speech and other vocalizations) from the audience. Here, the impulsive "spike" of the stroke is clearly visible, although we did find other examples where a weak stoke was barely discernable amongst the background noise either on the audio soundtrack or in the spectrogram. However, when they did not occur simultaneously, the spectrograms of strokes and the other types of sounds were clearly quite different.

5.2 Use of Templates of Mel Frequency Cepstral Coefficients (MFCCs)

The Mel Frequency Cepstrum[4] is a technique widely used in speech processing and is an alternative to the Fast Fourier Transform (FFT). The Mel cepstrum involves taking the Discrete Cosine Transform of the logarithm of the output of a filterbank, where each of the filter bands are band-pass filters with a triangular frequency response. The bands are typically not of equal width. (The logarithm is used to compress the range of possible amplitudes of the filter outputs.) Two of its main benefits are its strong ability to pick out the dominant periodic component from a signal, and its similarity to the perceptual response of human hearing.

Following a standard procedure used in speech processing, we have used 13 Mel Frequency Cepstral Coefficients (MFCCs) – 12 for frequency bands covering the range 0 – 15 kHz and one representing the power in the audio signal. For each MFCC calculation, we use a window of duration 10 ms.

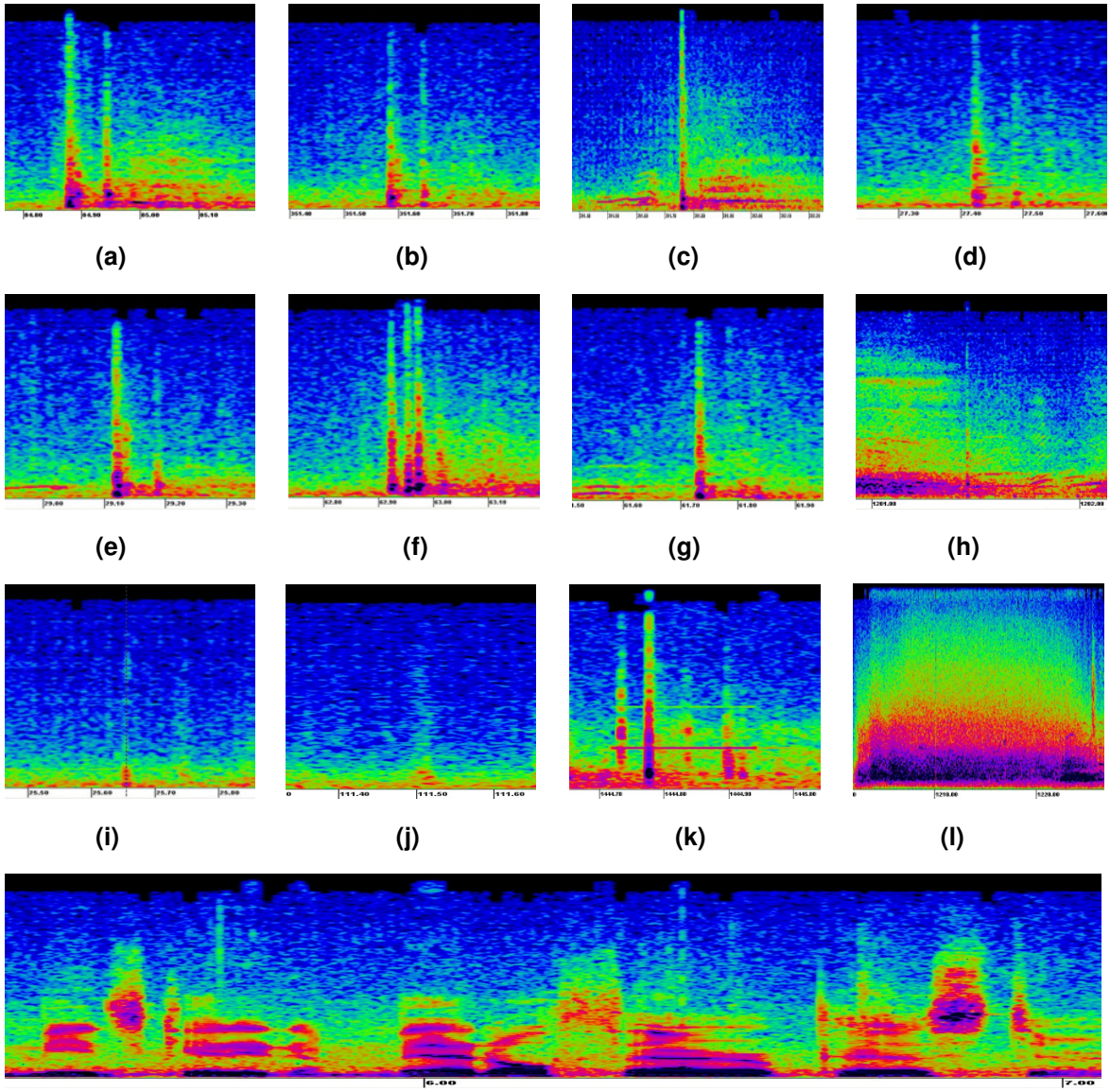

Figure 2 : Examples of spectrograms of various types of sounds occurring in tennis matches.
(a) First serve, (b) Forehand drive, (c) Forehand drive with “grunt” from the player, (d) Backhand drive (one-handed), (e) Two-handed background drive, (f) Smash (including sound from footstep), (g) Lob (with some speech in background), (h) Weak lob in a lot of audience noise, (i) Bounce of ball, (j) Footstep, (k) Ball clipping the net (with Cyclops “bleep”), (l) Applause from the audience, (m) Speech from a commentator

We then form a “template” of 20 such MFCC values, calculated over 20 windows, with successive windows overlapping by 5 ms, to create what is in effect a “slowly sliding template window”. (Any given MFCC window covers a time period which overlaps by 5 ms with its predecessor and by 5ms with its successor in the template). Each template thus represents sound events occurring within a period lasting 100ms in total. This value was chosen since it is relatively long compared with the duration of “impulsive” sound events (such as racket on ball impacts and bounces of the ball) occurring during the tennis matches, but short compared with the typical intervals (of the order of 1 second) between such events.

6 AUTOMATED DETECTION AND IDENTIFICATION OF SOUND EVENTS USING *PCA*

Our method of "encoding" the sound events in templates of MFCC values results in feature vectors of rather large dimensionality (20 x 13 = 260 in this case) for each example. In order to perform classification of examples into the 14 categories of sound events, it is highly desirable to work with a smaller number of "most useful" features. Principal Components Analysis[17,18,19] (PCA) is one method for achieving this, by choosing the "best" linear combinations of the original features. This can be likened to choosing the optimal angles to view an N-dimensional scatter plot of the N-dimensional data (for N small, e.g. N = 3) so that the different categories appear "best separated".

PCA proceeds by computing the eigenvalues and corresponding eigenvectors of the covariance matrix of all the examples for each category or class in the dataset of interest. In general, for data in N dimensions, there will be N eigenvalue-eigenvector pairs (not necessarily all distinct, but all the eigenvalues are non-negative) . In PCA, the largest M eigenvalues (for M < N) are retained along with their corresponding eigenvectors. These will form a basis for a "reduced dimensional" space which will be the "most useful M dimensions" for identifying data of that category. The choice of M can be empirical in order to obtain satisfactory classification performance. In application, data points (templates in this case) in the original space are projected onto the reduced (M dimensional) eigenspace for each category in turn. The example is put into whichever category gives the smallest distance (using an appropriate metric) between the original data point and its projection.

Here, each template is regarded as a feature vector of N = 20 x 13 = 260 components. Having used the video footage to identify the appropriate category (9 different types of tennis strokes and 5 "other events" – echo, ball bounce, applause, speech or silence) for each "sound event", the covariance matrix over all examples of that type was found. All the eigenvalue-eigenvector pairs for each category/matrix were then computed and the appropriate reduced eigenbasis for each class found by PCA. For various values of M << 260, we studied the effects on the successful classification rates of retaining only those dimensions in the eigenbasis corresponding to the largest M eigenvalues. In application, we used a Euclidean distance metric for determining the "closest" category (in the feature space) to any given example sound event, and that example was then classified as belonging to that category.

	Total	Num. Rec.	% Rec	Dimensions	% eigenvalue sum
[S1] First Serve	150	150	100	58	95.07
[S2] Second Serve	64	62	96.9	38	95.29
[FD] Forehand Drive	257	252	98.1	69	95.16
[BD] Backhand Drive	146	146	100	69	98.09
[2HBD] Two-handed BD	110	107	97.3	50	95.18
[SM] Smash	5	5	100	10	100
[VO] Volley	3	3	100	10	100
[SS] Stop Shots	8	8	100	10	100
[LO] Lob	7	7	100	10	100
[EC] Echo	298	283	95	62	95.5
[BC] Bounce	216	213	98.6	45	98.87
[SL] Silence	50	50	100	22	99.47
[AP] Applause	64	64	100	49	95.17
[SP] Speech	126	120	95.2	62	98.03

Table 3 : Statistics of Classification Success Rates for the various classes of sound events. The columns are : Type of sound event, Total number in dataset, Number of those recognised correctly, Percentage recognised correctly, Number of dimensions retained after PCA, Percentage of the eigenvalue sum due to those dimensions.

7 RESULTS AND DISCUSSION

Our initial results of applying PCA to templates of MFCC coefficients for the 14 categories of sounds we have specified for our data are very encouraging. Unlike some previous studies in the analysis of sports audio (e.g. Zhang & Ellis' work on basketball[5]) we have achieved very high correct

detection and classification rates for all 14 classes (see Table 3). Of the original 260 feature dimensions, we had to retain between 10 and 70 eigenvalues/vectors in order to correctly classify at least 95% of examples in any given category. However, it should be noted that all the cases where a very small number of eigenvalues/eigenvectors seemed sufficient were categories containing very small numbers of examples, and these observations may be indicative of "overtraining" the system.

The "confusion matrix" showing the number of examples of each type i classified as being of type j (with the cases where i = j corresponding to correct classifications) is shown in Table 4. It can be seen that, generally, the great majority of strokes were classified correctly, but a few second serves were mis-classified as forehand drives, one forehand drive classified as a backhand drive and one double-handed drive as an ordinary (one-handed) backhand drive. These pairs of strokes where confusions arose are, in all cases, of similar power and spectrographic profile. It is unlikely that the human eye could reliably distinguish between the two strokes of these pairs by visual inspection of their spectrograms (see examples in Figure 2(a) to (e)). Our system also performed well on classifying the sounds which were not of the strokes themselves. There were a few cases where bounces of the ball were mis-classified as echoes, and vice-versa, and a few (presumably weak) bounces and echoes were identified as "periods of silence". More surprisingly, a few examples of speech were mis-classified as echoes or bounces. These cases are being further investigated.

	S1	S2	FD	BD	2HBD	SM	VO	SS	LO	EC	BC	SL	AP	SP
S1	150	0	0	0	0	0	0	0	0	0	0	0	0	0
S2	0	62	2	0	0	0	0	0	0	0	0	0	0	0
FD	0	0	252	1	0	0	0	0	0	2	2	0	0	0
BD	0	0	0	146	0	0	0	0	0	0	0	0	0	0
2HBD	0	0	2	1	107	0	0	0	0	0	0	0	0	0
SM	0	0	0	0	0	5	0	0	0	0	0	0	0	0
VO	0	0	0	0	0	0	3	0	0	0	0	0	0	0
SS	0	0	0	0	0	0	0	8	0	0	0	0	0	0
LO	0	0	0	0	0	0	0	0	7	0	0	0	0	0
EC	0	0	0	0	0	0	0	0	0	283	12	3	0	0
BC	0	0	0	0	0	0	0	0	0	1	213	2	0	0
SL	0	0	0	0	0	0	0	0	0	0	0	50	0	0
AP	0	0	0	0	0	0	0	0	0	0	0	0	64	0
SP	0	0	0	0	0	0	0	0	0	2	4	0	0	120

Table 4 : "Confusion Matrix" for the Classifications and mis-classifications of the various sound event categories. The diagonal elements represent successful correct classifications. Row i column j represent the number of events of class i classified as being of class j.

On the whole, however, the high success rate of our system on this set of data is very encouraging. We are studying whether it is "robust" when applied to other tennis data and intend to investigate whether it could be used in other situations, which may be less controlled acoustic environments.

8 CONCLUSIONS AND FUTURE WORK

We have shown how the use of some techniques which have been inspired by well-established methods from automatic speech recognition can, in conjunction with Principal Components Analysis, lead to highly successful detection and classification of "salient sound events" which occur during the course of tennis matches during the Wimbledon Championships. Our methods were even able to distinguish between different tennis strokes very well, a task which would be very difficult for a person to do based on visual inspection of the appropriate spectrograms. We are currently working on using both discrete time (event orientated) Markov models and continuous time analysis (based on the time intervals between sound events) to look at sequences of these events and give our model predictive capability in addition to its current use in analysing the sound signal. Again, this is analogous to automatic speech recognition, where a "language model" is used to predict whether a proposed sequence of phonemes, syllables or words is a *probable* or *improbable* sequence, based on "past experience".

We also plan to extend this work to other situations : sports such as cricket, the detection of salient sounds in surveillance & security applications, or to monitoring the well-being of the sick, infirm or elderly. However, we note that many of these cases are less controlled environments than the tennis case, so their analysis and modelling will present new challenges.

9 ACKNOWLEDGEMENTS

This work has been funded by the EPSRC of the U.K. as part of the CAPS project (Grant GR/S78841/01). Krzysztof Zienowicz is grateful to the EPSRC for financial support.

10 REFERENCES

1. H. Brody, R. Cross & C. Lindsey, *The Physics & Technology of Tennis*, Racquet Tech (USRSA) Publishing, Solana Beach, U.S.A. (2002)
2. S. Hughes, The "Snickometer", http://experts.about.com/e/s/sn/Snickometer.htm (2005)
3. S. Rosen & P. Howell, *Signals and Systems for Speech and Hearing*, Academic Press, London (1990)
4. J. Holmes & W. Holmes, 'Speech Synthesis and Recognition', 2nd edition, Taylor & Francis, London, U.K. (2001)
5. D. Zhang & D. Ellis, Detecting Sound Events in Basketball Video Archive*, Technical Report, Electrical Engineering Department,* Columbia University*,* U.S.A. (2001).
6. R. Dahyot, A. Kokaram., N. Rea & H. Denman, Joint Audio-Visual Retrieval for Tennis Broadcasts", *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP '03)* (2003)
7. W.H-M. Hsu, Golf Impact Detection with Audio Clues, *Speech and Audio Processing and Recognition Project*, http://www.ee.columbia.edu/~winston/courses/speechaudio/project/E6820PrjRpt.pdf (2002)
8. Y. Rui, A. Gupta & A. Acero, Automatically extracting highlights for TV Baseball programs, *Proceedings of ACM Multimedia 2000*, Los Angeles, CA, pp. 105-115 (2000)
9. B. Zhang, W. Dou & L. Chen, Ball Hit Detection in Table Tennis Games Based on Audio Analysis, *18th International Conference on Pattern Recognition (ICPR'06)*, pp. 220-223, (2006)
10. W. Lao, J. Han, & P.H.N. de With, Automatic Sports Video Analysis using Audio Clues and Context Knowledge, *Proceedings of EuroIMSA* 2006, pp. 198-202 (2006)
11. A. Dufaux, L. Besacier, M. Ansorge, F. Pellandini, Automatic Sound Detection and Recognition for Noisy Environment," *Proceedings of the X European Signal Processing Conference* (2000).
12. D. Hoiem, Y. Ke, and R. Sukthankar, SOLAR: Sound Object Localization and Retrieval in Complex Audio Environments, *Proceedings of ICASSP* (2005)
13. T. Zhang and C. Kuo, Content-Based Classification and Retrieval of Audio", *SPIE Conf. on Adv. Signal Processing Algorithms, Architecture and Implementations VIII* (1998)
14. T. Zhang & C. Kuo, Hierarchical System for Content-Based Audio Classification and Retrieval, *Proceedings of the International Conference on Acoustics, Speech & Signal Processing (ICASSP)*, (1999)
15. Hawk-Eye Innovations, Instant Replay Comes to Tennis, http://www.hawkeyeinnovations.co.uk (2006)
16. G. Hunter, A. Shihab & K. Zienowicz, Modelling Tennis Rallies Using Information from both Audio and Video Signals, *Proceedings of the I.M.A. International Conference on Mathematics in Sport*, pp. 103-108, Salford, Manchester, U.K. (June 2007)
17. R. O. Duda, P. E. Hart & D. G. Stork, *Pattern Classification*, 2nd Edition,Wiley, New York, USA, pp 568 (2001)
18. I.T. Jolliffe, *Principal Component Analysis*, 2nd Edition, Springer, New York, USA (2002)
19. M. Kendall, A. Stuart & J.K. Ord, The Advanced Theory of Statistics, Volume 3 (4th Edition), pp 320-369 (1983)