THIRD INTERNATIONAL CONFERENCE ON IMAGE PROCESSING AND ITS APPLICATIONS
University of Warwick 18-20 July 1989
IEE Conf Pub 307, pp 300-304

**THE RECOVERY OF 'PHONOVISION'**

D F McLean

## INTRODUCTION

The history of the recording of moving television images has its beginnings in the late 1920's. At that time, John Logie Baird and others sought to implement a practical television system by scanning the scene opto-mechanically, initially using a variant of the Nipkow disc (1). Although eventually superseded by developments in electron tube technology, Baird demonstrated many of the principles of modern television - including the world's first TV recordings.

## BACKGROUND TO 'PHONOVISION'

From 1926 to 1928, Baird invented methods of recording and reproducing vision and sound signals onto wax cylinders (2) and discs. He subsequently called this process *Phonovision.* To date only five different Phonovision discs have been discovered. These span the period September 1927 to March 1928.

Researching contemporary publications reveals that unlike Baird's other experiments, the reproduction of images from Phonovision discs was never demonstrated, although Phonovision itself was well publicised. Since then and up until the work described here and in McLean (3), no recognisable results have been reproduced from these discs. The failure to do so is a result of the poor quality of the recordings.

## PRELIMINARY ANALYSIS

The recordings comprise a heavily distorted video raster waveform. Each successive raster or frame is made up of 30 vertically-scanned lines consistent with the Baird 'standard' of 1926-1935. Notably there are no embedded synchronisation pulses identifying the start of frame or line.

90 lines (3 complete frames) were recorded on each revolution of the disc. Each disc holds approximately 750 frames. As the Baird standard had a frame repetition rate of 12.5 per second, this implies that the disc had to be played back at 250rpm with a total duration of approximately 60 seconds.

## OVERALL APPROACH

Knowledge of the distortion mechanisms in the complete process (imaging, recording and reproduction) would have allowed implementation of controlled optimal restoration. As no collateral data was available on the material or the distortion mechanisms, the procedure adopted was to attempt correction by iterative measures with the objective of improving image quality. Measures of the specific correction applied were then analysed in an attempt to quantify the appropriate distortion mechanisms.

Initial analysis of the imagery indicated two basic types of distortion:
- distortion of the vision signal
- picture synchronisation problems

## DISTORTION OF THE VISION SIGNAL

The maximum bandwidth occupied by the studio video signal prior to recording was from DC to about 13kHz assuming equal horizontal and vertical resolution. The observed bandwidth on playback is equivalent to 800Hz to 8kHz with phase and frequency response variations in the pass-band. This distortion is likely to have occurred in the recording process.

The video waveform was electromechanically recorded onto the surface of the wax disc as a base-band signal. From the consistent 'WT' suffix in the disc serial numbers, it is assumed (4) that the Western Telegraph process, the most modern available in the 1920's, was used. In the absence of a specification, it is believed (4) that the pass-band frequency response was approximately 300Hz to 9kHz and was not 'flat'. It is also understood that the quality of recordings made with this process was variable and dependent mostly on the mechanical coupling of the head-cutter with the disc surface and the temperature (and hence plasticity) of the wax.

The distortion can be classified into the following categories:-
a) limited pass-band with respect to the video signal bandwidth
b) time-variant pass-band frequency and phase non-linearities
c) masking of low frequencies by uncorrelated noise ('rumble') present in the recording.
d) high frequency surface noise
e) resonance of the head cutter

### Suppression of Signal Distortion

For distortion to the frequency response, in categories a) and b) above, it was not possible to provide deterministic restoration of the video signal without quantitative knowledge of the distortion. As a result, attempts at Fourier domain filtering (5) by iterative means were unsuccessful. With the facilities available, this computationally intensive approach was abandoned in favour of an analogue filter bank and phase correcter manually adjusted for best subjective improvement.

Low frequency 'rumble', category c) above, extends up to 200Hz and is probably caused by vibration of the recording apparatus. It was removed by convolving the data with a one-dimensional equal coefficient high-pass filter. The results can be seen in Figure 1.

The uncorrelated high frequency surface noise of category d) distortion was suppressed in a different way, based on the low information content of the image both spatially and temporally. The approach used was to filter the image temporally by exploiting the high correlation between adjacent frames. The low frame repetition rate coupled with occasional fast subject movement restricted filtering to two consecutive frames. The subject movement from frame to frame meant that the temporal filtering had to track and compensate for this motion without degrading the image.

The motion compensation technique applied here considers a 3 by 3 spatial region around each pixel to be filtered. Assume this pixel is at sample position, i, in line, j, in frame, k, and whose brightness is P(i,j,k). The sum of differences, SD, is calculated for this spatial region and the corresponding region in the (k+1)th frame.

$$SD(x,y) = \sum_a \sum_b ( P(i+a+y, j+b+x, k+1) - P(i+a, j+b, k) ) \quad \textbf{(1)}$$

where a and b have values -1,0,1 and x,y define the spatial offset for the search within the subsequent, (k+1), frame. (Recall that samples increment vertically (in y) and lines increment horizontally (in x)).

For each value of x and y, this expression gives an approximation to the correlation between regions in adjacent frames. The search range of x,y in the next frame was limited to a 5 by 3 sample region, based on an analysis of typical subject movement.

Once the highest correlation is found at, say x',y', the centre values of each region are averaged in the following way:-

$$P(i,j,k) = ( P(i,j,k) + P(i+y', j+x', k+1) )/2 \quad \textbf{(2)}$$

The result of applying this technique is illustrated on Figure 2.

Head-cutter resonance, category e) distortion, occurs only occasionally on the Phonovision recordings and only for short periods. It does occur continuously however throughout the 'Major Radiovision' recording made eight years after the Phonovision recordings and just prior to the demise of the 30-line broadcasts. In this case the resonance was removed by a Q=10 digital notch filter. The result is shown on Figure 3.

## PICTURE SYNCHRONISATION PROBLEMS

Keeping the display mechanism synchronised with the 30 line picture was a persistent problem right through to the demise of the 30 line service in 1935. For recorded television, Baird circumvented the problem of synchronisation on playback with the invention (6) of an ingenious device called a 'Phonovisor'.

The commercial success of such a device was entirely dependent on the ability to record the vision signal with the scanning disc in the camera synchronised and geared down to provide exactly three TV frames (ref 3) on each revolution of the record. The analysis here shows that the recording process was not without serious problems.

### Classification of Timebase Irregularities

The video data was sampled at a constant rate using a clock synchronous with the disc platter rotation. The resultant imagery shows signs of variation in the original recording speed. This variation can be classified into the following categories:-

Variation in line position in every frame. Here the pattern within frames is invariant throughout the recording. This is caused by poor construction of the Nipkow scanning disc. The error is consistent with the incorrect angular positioning of the scanning apertures. In the 30-line Nipkow disc, the scene was scanned by 30 successive apertures spaced at exactly equal angles around the disc periphery. An angular positioning error would cause some lines to be scanned slightly early or late. The maximum error observed is 3% of line length and is equivalent to a 0.4 degree error in circumferential position of that aperture. Interestingly this pattern can be observed on all early recordings indicating that the same scanning disc was in use.

High frequency variation whose within-frame pattern develops throughout the recording. This is present on only one recording, on which the pattern alters and develops within the first few seconds of the recording and then stabilises. As the pattern is at the frame rate, the deduction is that it is a high frequency resonance induced by mechanical vibration in the coupling between the Nipkow scanning disc and the Phonovision apparatus.

Low frequency variation. Present on all recordings, this extends in frequency from a long-term drift to variations at the frame repetition rate. The effect is at its worst on the September 1927 recording, is much reduced on those of January 1928 and is yet again much reduced on the latest studied, recorded in March 1928. This infers that improvements were being made to the apparatus to counter this effect.

The March 1928 recording shows a long-term drift in speed of 0.005%. Such precision implies a mechanical link with the scanning disc. The drift was possibly caused by using a belt drive, clutch or flexible coupling.

### Approach to Restoring Synchronisation

The absence of any synchronisation information meant that correction of this distortion had to be based on the image content. For assistance in the removal of this distortion, an automatic correlation matching and tracking technique was developed.

Given a template, T(x), corresponding to the first frame in a sequence, the next frame, A(x), at offset, s, is found by looking for a maximum likelihood of occurrence using correlation matching. That is,

$$CS(s) = \sum_x ( T(x) * A(x+f-s) ) \quad \textbf{(3)}$$

where x spans the no. of samples within a frame, and f is the index to the start of the frame.

The range of values for s is set at -64 to +63 since, in the worst case, the variation does not exceed 64 samples (one line) per frame.

When the maximum value of CS(s) is found, maxCS(s), the offset corresponding to this value is added to the current frame pointer, f, to create a pointer to the start of the new frame, f'. The new frame pointer, f', is stored in a list together with a normalised version of maxCS(s). Normalisation to the average signal amplitude in the frame reduces the dependency of the correlation score on signal amplitude.

At this point, the contents of the template - initially set to the values of the first frame in the sequence - are integrated with the new frame data.

$$T(x) = (1-K(maxCS))*T(x) + K(maxCS)*A(x+f') \quad \text{for all x within one frame} \qquad \textbf{(4)}$$

where K is a variable in the range 0 to 0.5 and is a function of the maximum correlation score, maxCS.

A high value of maxCS results in K being set to its maximum value of 0.5. A correspondingly low value sets K to 0 disabling the update of the template. This 'leaky bucket' integration technique serves to improve the stability of the matching and tracking process.

Post-Processing

After a sequence had been processed, the frames were corrected for their mean position. To understand the need for this, consider the effect of a drop of 1% in the recording speed. The image data would be temporarily over-sampled and the image would shear at approximately a 45 degree angle.

To compensate for this effect, each frame and its neighbour were further processed by applying a fraction of the offset linearly interpolated between the centres of the adjacent frames.

Results of Correlation Tracking of Frames

The technique above was developed progressively with experimental verification. Nevertheless the technique could not be made stable in tracking sequences in excess of 60 frames. Its residual jitter was poor from frame to frame mainly because it can only correct for recording speed variation at up to the frame rate. However, as can be seen in Figure 4, components of speed variation exist right up to the line frequency.

The dependency of the algorithm on video content meant that the technique was unstable when the scene changed radically (especially on disappearance of the subject). Manual intervention was necessary in these instances and for accurate alignment of sequences.

## Spectral Analysis of the Time Domain Distortion

Part of a recording shows a horizontal white bar. A line-by-line correlation technique was adapted from the frame method above and the offsets for 512 sequential lines were stored, de-trended and Hamming windowed prior to calculating the power spectrum (Figure 4). The dominant peaks are at the rotational period of the Phonovision disc (90 lines period) and at its 2nd harmonic phase-shifted by 180 degrees. Although this could be caused by misalignment of the centre of disc rotation (only a 0.5mm offset equates to a 12% line-time fluctuation with a 3 frame period), the presence of 2nd harmonic infers strongly either an imbalance or parallel or angular misalignment of the linkage. Other peaks occur at the frame rate and its 2nd and 4th harmonics. These arise from an incorrect assumption that the white bar is perfectly horizontal.

## GEOMETRIC DISTORTION CAUSED BY ARC-SCANNING

The use of a Nipkow disc camera in the creation of Phonovision imagery can be proven through analysis of the September 1927 recording. For part of this recording a ventriloquist's dummy head is moved from one side of the TV frame to the other.

Measurements accurate to 0.8% were made of the distance along the medial axis of the dummy's face for all lateral positions in the short sequence of the recording. The resultant graph, together with a least squares fit, (Figure 6) shows the object height (expressed as a percentage of line length) plotted against the line number on which the measurement was taken. A trend of height varying with line number indicates that the image was scanned by a Nipkow disc.

Further, as the fractional height is larger near line 1, this indicates that the first line in the frame was innermost and the last line was outermost. This is the opposite of the Baird TV 'standard' used in transmitting between 1929 and 1935 and in the 1935 'Major Radiovision' recording but is in agreement with photographs taken in 1928 (7)(8).

## Derivation of the Image Aspect Ratio

From the graph in Figure 6 and the assumption that the Nipkow scanning disc comprised one frame of 30 apertures, the aspect ratio can be derived.

Referring to Figure 5, ABCD represents the scanned area on the periphery of the Nipkow scanning disc with angle AOB being 2*PI/30. Arcs AB and DC represent the paths that line 1 and line 30 follow in the raster (from the previous section).

As the angle here is small (about 4 degrees) the object heights at lines 1 and 30 are:-

$$x/R1 = \sin f \; ; \; y/R30 = \sin g \qquad \textbf{(5)}$$

where R1 and R30 are the radii to lines 1 and 30 respectively,
f = angle POQ = F1*2*PI/30
and g = angle SOR = F30*2*PI/30

F1 and F30 are the fractions of line length for lines 1 and 30 respectively. They are calculated using the equation (on Figure 6) for the best fit graph.

As the object's height is constant across the frame (x=y), the equations can be combined to derive an expression for the <u>width</u>, W, of the TV frame:

$$W = R30 - R1 = R1*(a - 1) = -R30*(1/a - 1) \quad \text{where } a = \sin f / \sin g \qquad \textbf{(6)}$$

Now, the raster height can be expressed by the arc-length, H.

$$Hx = Rx*2*PI/30 \text{ ;where x is line number} \qquad \textbf{(7)}$$

The aspect ratio is defined as the ratio of the height to width of the scanned area, where the height in this case is taken to be at the centre of the display. The instantaneous aspect ratio at the extremities of the display are :-

$$AR1 = H1/W = -2*PI/( 30*(1/a - 1) ) \quad \textbf{(8)}$$

$$AR30 = H30/W = 2*PI/( 30*(a - 1) ) \quad \textbf{(9)}$$

Taking the average of these two values gives an expression for the aspect ratio of the display, AR.

$$AR = -PI*( (1+a)/(1-a) )/30 \quad \textbf{(10)}$$

With F1 = 0.36037 and F30 = 0.32644 from the least squares fit (Figure 6) the calculated aspect ratio, AR, is **2.12** (height to width).

This value is within 10% of the value associated with the Baird 'standard' of 7 units high by 3 wide, or a ratio of 2.33.

## CONCLUSIONS

This paper has described the procedures used in the enhancement and analysis of the earliest known recordings of television pictures.

Information not previously available has been derived analytically from the features and defects of the Phonovision recordings. The specific areas that are now better understood are:-
- the original image scanning process,
- the method of image recording,
- the apparatus used and
- the difficulties encountered.

From the material studied, the conclusion is that Baird was unable to proceed further with Phonovision. The concept was ahead of what technology could offer.

## ACKNOWLEDGEMENTS

The author is a Principal Consultant with Logica Space and Defence Systems Ltd and would like to thank them for permission to publish this article.

## REFERENCES

1 Nipkow P, 1884, DRP Patent No 30105
2 Baird, J L, British Patent No 289104 applied 15th October 1926
3 McLean D F, "Computer-based Analysis and Restoration of Baird 30-line TV Recordings", Journal of the RTS, April 1985
4 Cook, G. Columbia Graphophone Company, Private Communication
5 Andrews & Hunt, "Digital Image Restoration", Prentice-Hall, 1977
6 Baird, J L, British Patent No 324049 applied 10th October 1928
7 Moseley & Barton Chapple, 2nd edn 1931, "Television Today & Tomorrow", plate opp p56
8 Television, Nov 1929, pp448-451

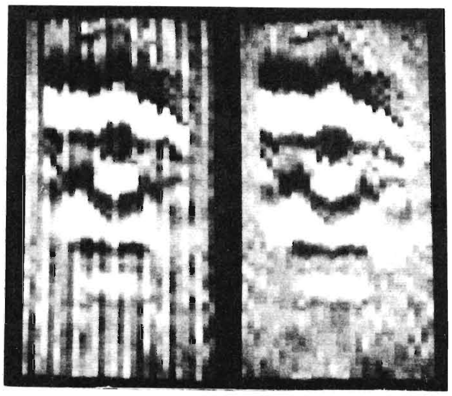

**Figure 1** Reduction the effect of disc 'rumble' by high pass filter. *(RWT 620-4 10th January 1928)*

**Figure 2** Suppression of noise by motion-compensated time domain averaging filter. *(RWT 115-3 28th March 1928)*

**Figure 3** Removal of resonance induced by head cutter during recording and correction for arc-scanning. *(Major Radiovision 1935)*

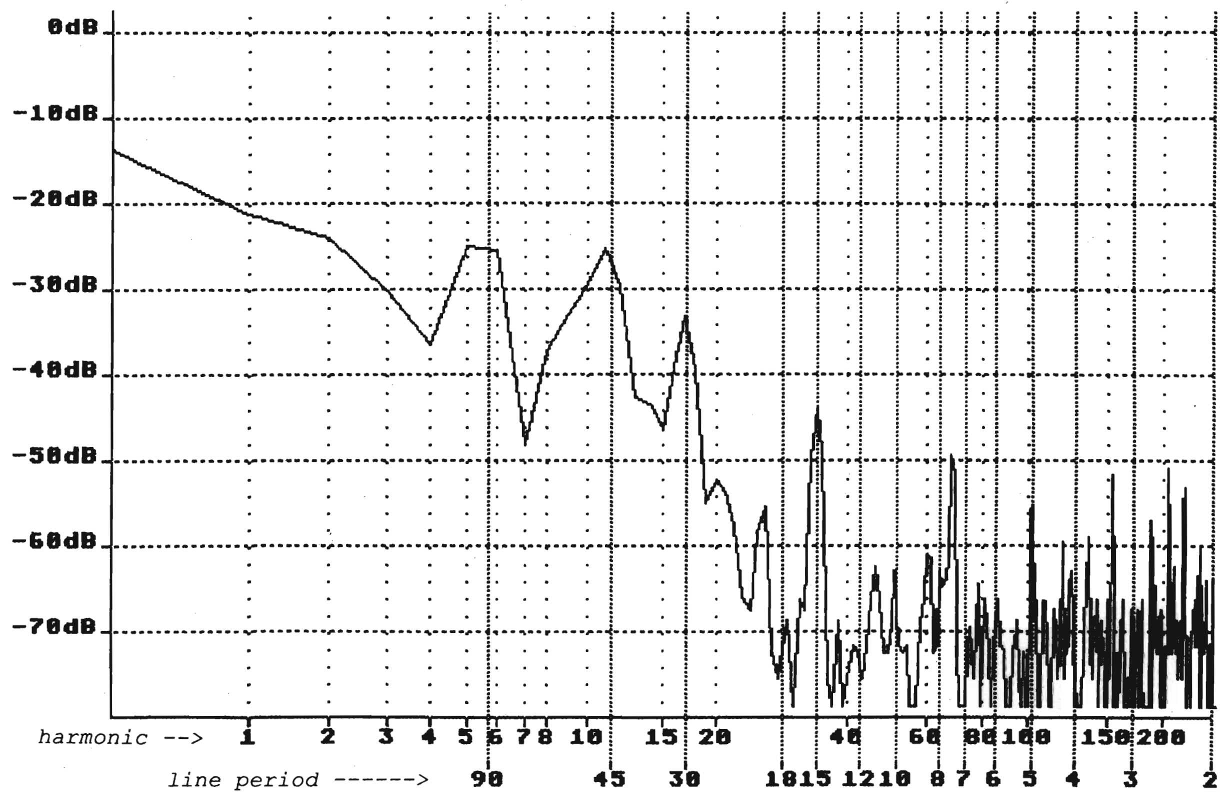


**Figure 4** Power Spectrum of Variation in Speed over 17 TV frames.
*(SWT 515-4 20th September 1927)*

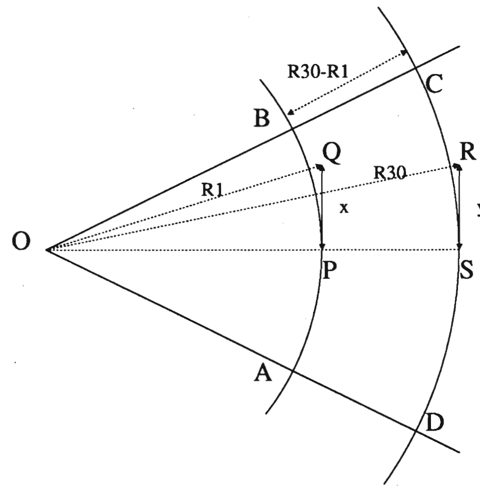


**Figure 5** Schematic Diagram of 30-line arc-scan raster on Nipkow disc

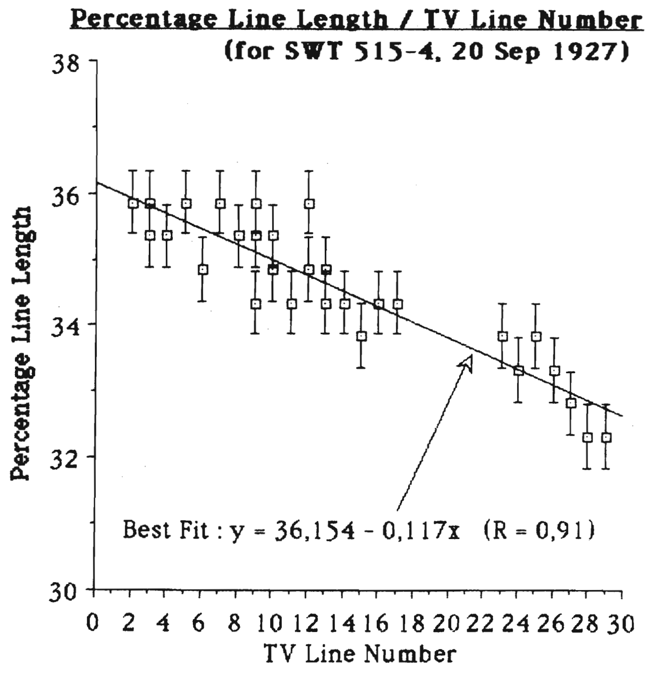


**Figure 6** Graph showing characteristic distortion of height as a function of line number in the TV raster.
*(SWT 515-4 20th September 1927)*