THE QUEEN'S AWARDS
FOR ENTERPRISE
2001

BBC

R&D White Paper

WHP 123

September 2005

Music Production Planning - Updating the Process After 50 Years!

M.J. Evans, D.G. Kirby *and* D. van Kemenade

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 123

M.J. Evans, D.G. Kirby, D. van Kemenade

Music Production Planning - Updating the Process After 50 Years!

Abstract

The pre-production camera scripting required for a live music production is extremely detailed and, using current methods, can take up to 12 hours to produce the script for a typical three-minute music item. We describe a new approach based on automatic music analysis coupled with a standard video editing system used, not to edit video, but, in effect, in reverse to produce the script required to shoot the video later in the studio.

Although experimental, the system has been used for a number of pop music productions and is beginning to revolutionise how the planning for such programmes is carried out.

This document was originally published in the Proceedings of the International Broadcasting Convention, September 2005.

MUSIC PRODUCTION PLANNING:
UPDATING THE PROCESS AFTER 50 YEARS!

M. J. Evans, D.G. Kirby, D. van Kemenade

BBC R&D, UK

ABSTRACT

The pre-production camera scripting required for a live music production is extremely detailed and, using current methods, can take up to 12 hours to produce the script for a typical three-minute music item. We describe a new approach based on automatic music analysis coupled with a standard video editing system used, not to edit video, but, in effect, in reverse to produce the script required to shoot the video later in the studio.

Although experimental, the system has been used for a number of pop music productions and is beginning to revolutionise how the planning for such programmes is carried out.

INTRODUCTION

With no opportunity for retakes or working around problems in post-production, shooting a live music performance relies on having a camera script that has been carefully planned beforehand. For a fast paced pop-music item lasting three to four minutes, there can be up to 200 camera shots, with each timed carefully to the beat of the music and the flow of the overall performance. Despite the intricate detail required from this planning process, the system has changed little over the past 50 years: it has remained a manual process relying on paper and pencil.

Although simple performances can be covered without such a detailed camera script, this requires that shots and cuts between cameras must be improvised ‘on-the-fly’. For a busy dance routine or other complex performance, even with the benefit of prior knowledge of the flow of the music, the results achieved in this way can be far from ideal.

The preferred approach that has evolved from the experience of both broadcasters and the music industry is to plan the camera shots and timings carefully so that they match both the music and choreography. This allows the sequence to be repeated in a consistent way, so that it can be improved during rehearsal, and the final, live performance should have few surprises.

There are three main stages in planning a music production in this way. Firstly, the script supervisor will carry out the initial music breakdown to annotate the music track. This will include the lyrics of the song, the beats and bars of the music, and how the lyrics are timed to the beats. It may also include comments indicating the general structure of the music, such as verses and choruses. For a three-minute music track, breaking the music down to the required detail in this way can take up to one hour. An example of this is shown in Figure 1. Here the lyrics are shown with comments and, around them, where lyrics are not being sung, the beats and bars are indicated by numbers.

Once this initial script of the music content is ready, it is passed to the TV Director who will plan the camera shots that are required. This is a task for paper and pencil! The shots, with their cameras details and description, will be written on the music breakdown, down to the level of individual beats of the music. This stage can take between two and six hours for a three-minute track but possibly more if there is intricate choreography.

The third stage of the planning process falls to the script supervisor who will take the Director’s marked-up script and type up all the details to produce the final script ready for the studio. In doing this, all the timing details need to be calculated, so that the duration of each shot in beats and bars can be included with its description. This final stage of preparing the script can take up to five hours for a three-minute track. Figure 7 shows an example from a completed script (although this example has been produced automatically, its appearance is identical to one created by hand). It is similar to Figure 1 but now includes the camera shots and timings.

Verse 2
Isn't where you wanna be
1 2 3 4
And isn't what you wanna do
1 2 3
Just give me
one more day
4
one more day
Give me another
2
night
4
just another night

Figure 1 – a music breakdown showing lyrics and beats

From this brief outline, it can be seen that the planning process is extremely time-consuming. For a typical music item lasting just three minutes when broadcast, the Director and assistant could have spent between 6 and 12 hours in planning the script prior to rehearsal in the studio. For a live broadcast, this scripting has to be accurate as there’s little time for changes once rehearsal is underway.

We have looked carefully at this planning process and the key stages that are required. From this we are developing a new way of planning which, so far, is offering considerable promise in terms of the time-savings that can be achieved.

NEW APPROACH

Our new approach adopts two of the three stages described above: music breakdown and shot planning, but automates these as much as possible. As described in more detail later, the music breakdown employs beat extraction to track the rhythm of the music. This is an essential starting point, as the majority of cuts between cameras will on a beat. The second step in the music breakdown is to produce the timings of the lyrics. Although the user can mark these manually, we are experimenting with using speech recognition techniques to automate this process too.

Once these two processing stages are complete, the timings for the track are available and an initial music script could be created. However, creating a printed script of the music is no longer the preferred way forward for the next step of planning the cameras shots. Instead, the shot planning moves to a video timeline editing system which allows the Director to visualise the intended shots and their timings much more easily. For this editing step, we are trying to avoid developing a bespoke timeline editor for planning but instead are exploring how off-the-shelf video-editing packages can be used to provide this functionality.

As there is no video available at this point in the planning process, we start by providing three audio tracks on the timeline: the music recording itself and two beat tracks, one a ‘click’ track and the second, a voice counting through each bar. Both of these latter two tracks are produced automatically from the music analysis processing.

With these audio tracks as a guide, the Director uses simple still images or graphics to represent the shots and inserts these onto the video tracks in the timeline to plan how the various cameras are to be used.

Once the Director has completed the planning in this way, all the information required for the script is available. The music breakdown stage provides the music timings whilst the edit list from the video editor provides the details of the cuts between the cameras and how they relate to the music. From this information, the final script to be used in the studio can be produced automatically by the software. In this way we eliminate completely, the final manual step of creating the script from the Director's handwritten notes.

The key stages in this process are described in more detail in the following sections.

MUSIC ANALYSIS

Requirements

In our work to date, we have deliberately restricted ourselves to working with pop music. Not only does this cover a significant part of the BBC's weekly music output on television but the nature of the music, generally with a strong percussive element, means that beat extraction should be more consistent. We have therefore concentrated on beat extraction from percussive music, although we intend to develop the music analysis algorithm further to cater for other genres of music.

Detecting and Tracking the Beat

Our music analysis algorithm starts in the same way as the method published by Scheirer (1), by detecting rapid changes in the audio envelope in several frequency bands. The resulting signals have high positive peaks at the positions of notes with sharp attack in the original music. Some of these percussive events will correspond to the beat of the music.

To detect the repetition rate of these events, rather than then using a resonant filterbank as in (1), our implementation, instead, divides the audio into short segments, each just less than three seconds long. The data from the seven frequency sub-bands is recombined to obtain the positions of all the percussive notes in each segment. Figures 2(a) and (b) show a segment of music and the percussive notes detected within it.

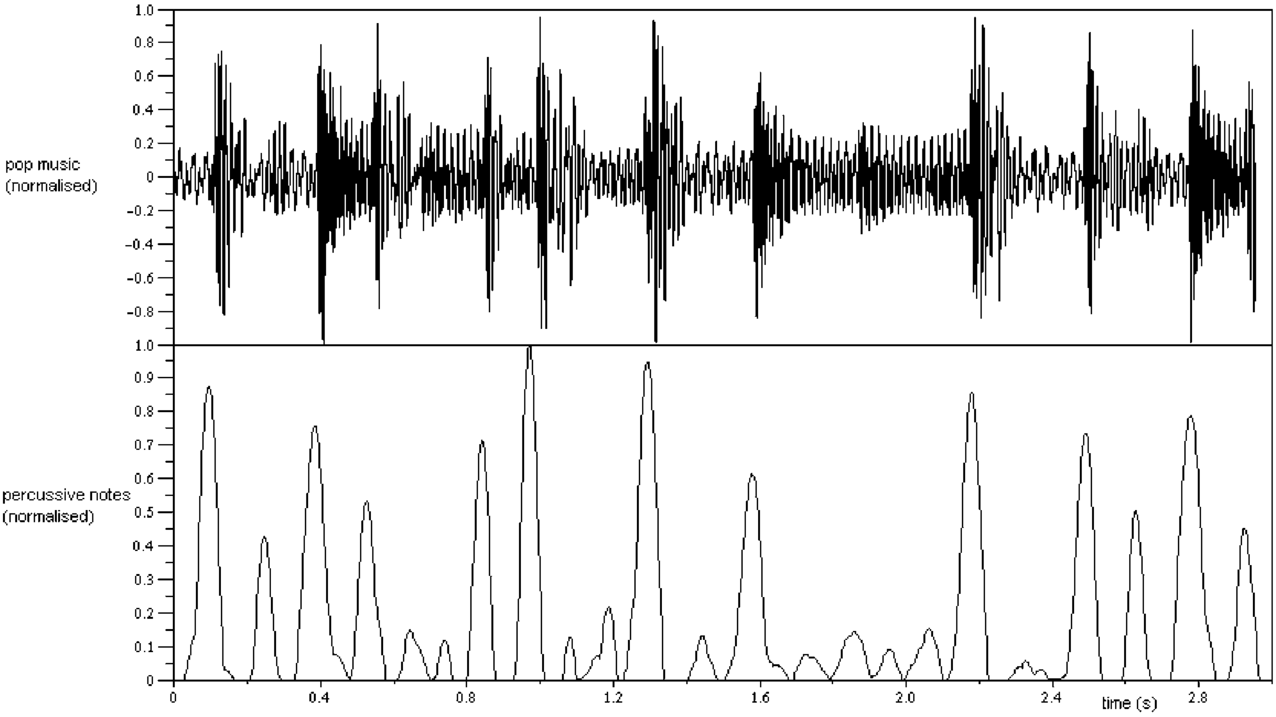


Figure 2 – (a) upper: The amplitude waveform of the segment being analysed
(b) lower: The positions of the percusive notes derived from sub-band analysis

We then apply the technique of autocorrelation, to identify percussive notes which are evenly spaced, as shown in Figure 3.

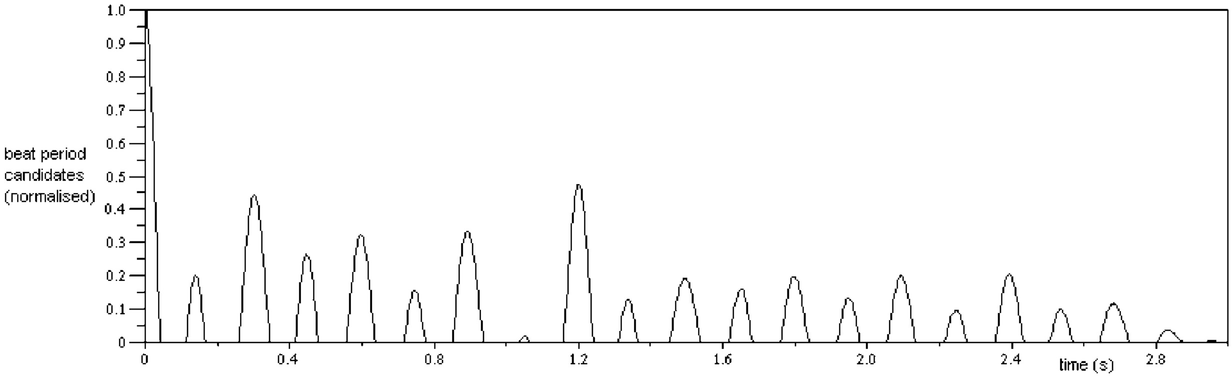
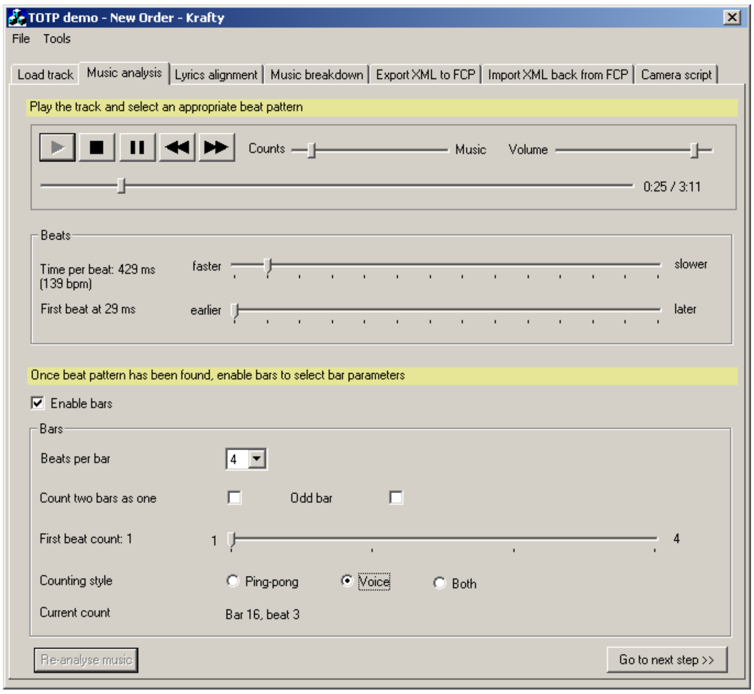


Figure 3 – Possible beat periods derived from Figure 2

Any even spacings are potential beat periods and the evenly spaced notes themselves are the candidate beats. There will always be several candidate beat patterns from the analysis.

Our experimental user interface for this processing is shown in Figure 4.

In using the music analysis system, the user starts the music analysis in a suitably rhythmic part of the track and chooses which of the candidate beat periods and phases for that segment is appropriate. For example, he/she might choose to mark the down- or up-beat, or have the beat counted at normal, or double speed. The beat pattern is then extended forwards and backwards across the remainder of the segments in the track. Automatic adjustments in beat frequency and phase are made at each segment boundary in order to create an accurate set of beat positions that track any drift in rhythm.



Play the track and select an appropriate beat pattern

0:25 / 3:11

Beats

Time per beat: 429 ms (139 bpm)

First beat at 29 ms

Once beat pattern has been found, enable bars to select bar parameters

☑ Enable bars

Bars

Beats per bar: 4

Count two bars as one □ Odd bar □

First beat count: 1

Counting style ○ Ping-pong ● Voice ○ Both

Current count: Bar 16, beat 3

Figure 4 – The user interface for the music analysis process

Beats and Bars

Although the algorithm can automatically detect the positions of beats in percussive music, it is a much more difficult task for the system to automatically identify the music's metre. However, for the majority of pop music, the user simply needs to specify the number of beats to a bar, and indicate the position of the first beat of one bar. The system then applies this division of the beats along the whole track, using the user's input as a reference.

Results

The music analysis algorithms have been tested in detail using around twenty tracks of current and archive pop music. This has allowed a fine-tuning of the system which has now been used to track the beats and bars of many more examples. On a typical PC, analysis takes around thirty seconds per three-minute track.

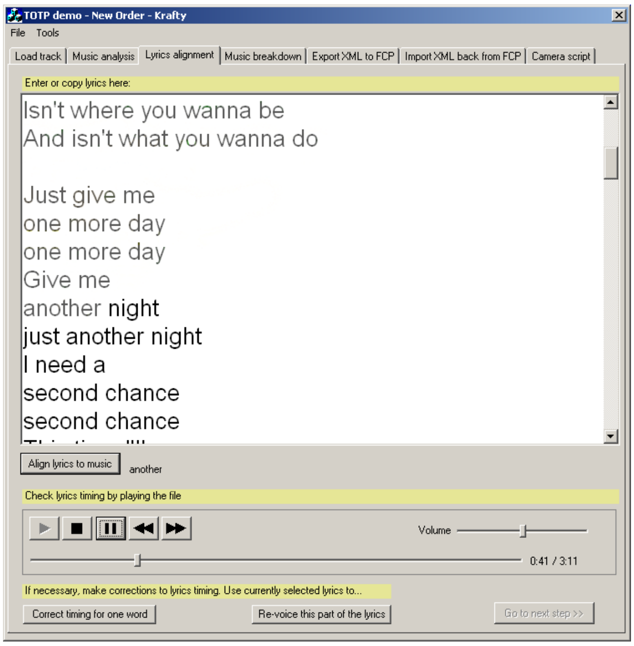
Currently, the grouping of beats into bars is fairly limited and is unable to cope with a change in the number of beats to the bar, although this is very uncommon. The occurrence of stray beats outside the bar structure is more common and also needs to be catered for. For accuracy, the algorithm is deliberately limited to only modest and gradual changes in beat period. This is not a serious limitation as it is extremely rare for a pop music track to contain gross changes in beat period. However, many pop tracks have a 'bridge' of contrasting material in their second half, and this can be non-percussive. On encountering such sections of music, the algorithm estimates beats which continue the prevailing beat pattern. Any limitations in this, or any other part of the system, have also been helped immensely by giving users the facility to 'pick-up' the beat analysis; restarting the detection from the segment in which it went wrong.

ALIGNING THE LYRICS TO THE MUSIC

Now that we know the timings for the beats, we also need to find out when each word of the lyrics is sung. In the old scripting process, the words were marked with beat numbers manually.

While looking for a way to automate this step, we considered our past work for subtitling (2) in which a speaker-independent speech recogniser is used to match the spoken words, taken from a programme script, with the programme soundtrack. This is a speech recognition process referred to as ‘alignment’.

Although the alignment software is intended for use with speech only, it turns out that it can often work with pop music as well. Despite the instrumental backing of the track, the alignment can still provide accurate estimates of the time at which each word is sung. The main limitation is that the track needs to have clear vocals. When there are multiple singers at any one time or a singer has been overdubbed, the alignment will get confused and the estimated lyric timings are no longer reliable. In those cases, our application’s user interface allows the timings of key words to be input manually, so that reliable automatic alignment of the other words may become possible. If this does not give reliable timings either, it is still possible to re-voice those parts of the lyrics that cause difficulties for the alignment. In this case, the user listens to the original track and speaks the lyrics into a microphone while they are being sung. The result of this is an alternative track with clean vocals that generally gives reliable word timings after alignment.



Enter or copy lyrics here:

Isn't where you wanna be
And isn't what you wanna do

Just give me
one more day
one more day
Give me
another night
just another night
I need a
second chance
second chance

another

Check lyrics timing by playing the file

If necessary, make corrections to lyrics timing. Use currently selected lyrics to...

Figure 5 –The lyrics display for entering the lyrics and matching them to the music

Figure 5 shows the lyrics display of our experimental system. This window allows the user to check and adjust the timings of the lyrics, with the text changing colour as the music is replayed.

We plan to improve this part of the process by experimenting with more advanced speech aligners and adding new features to the user interface that make it easier to input word timings manually.

SHOT PLANNING

With the music breakdown completed, the planning process moves to the programme

Director who must decide on the various camera shots and their timings. We are using a standard video-editing package for this, as there are considerable benefits to be gained if users are already familiar with particular commercial products. As far as possible we are trying to ensure that our approach will make it possible for any of the commonly used editing packages to fulfil this role. However, as we are in effect, using a video-editor 'in reverse' to create a camera script from the timeline rather than edit existing video content, the merits of the various packages are different in this context.

The key idea behind this part of the planning process is to use still images or graphics to represent camera shots. For example, a jpeg image containing the letters 'CU' can represent "close-up", whilst 'GS' can mean "group shot", etc. However, the actual content of the image is irrelevant as long as the Director understands which type of shot it represents. Simple graphics can also be used, for example, with a group shot being represented by a group of cartoon-like figures. Single video frames grabbed from a similar performance can also be used, if a more realistic image is preferred.

Planning using the timeline

The Director starts by opening the video project file that was created automatically by the music breakdown process. This provides the starting point for the planning by displaying three audio tracks and one video track on the editing timeline. The first audio track is the music recording itself. The second is a click track which indicates the beat and bar positions in the music with short 'chirps'. The third is similar but with a spoken voice counting the beats and bars, e.g. "1, 2, 3, 4", as a musician would.

The video track displays the lyrics which appear as subtitles, synchronised to the song as it is replayed. The timings for these words are produced at the earlier music breakdown stage, but the sequence of text images that form the video track is generated automatically by a "text generator" source within the video editor software.

As this point, when the timeline is played, the music will be heard with the click and/or counting track and the lyrics will appear overlaid in the picture as the words are sung. The Director can now develop the camera shots within this framework by creating additional, blank video tracks for each of the cameras that will be used.

Using the still images, previously prepared or from a library, the Director assigns shots to the cameras by dropping the corresponding image onto the appropriate camera track and extending its duration and position, as necessary. Subsequent shots can be added easily in the same way to build up the required sequence of shots across all the cameras. A 'snap-to-edit' or 'snap-to-marker' facility in the video editor becomes very powerful here, as timing markers, produced from the beat analysis, can be used to snap the cuts between cameras automatically to the beat of the music.

Although the same shot image, for example "CU", would be used each time a close-up shot is required, comments can also be added to each separate occurrence on the timeline, so that specific camera directions can be appended each time. Depending on the capabilities of the editing package being used, this provides a convenient way for the Director to add the necessary description to each shot for the cameraman.

Figure 6 shows a few shots of a performance built up in this way. In this case the Director has used simple graphic images to represent the various types of shot.

As the timeline gradually builds up with shots, it can be replayed to judge the timing of the shots and how they match the flow of the music.

In some cases, particularly where choreography is involved, a video recording of the dancers rehearsing may also be available. If this is laid down as a base video track in the timeline,

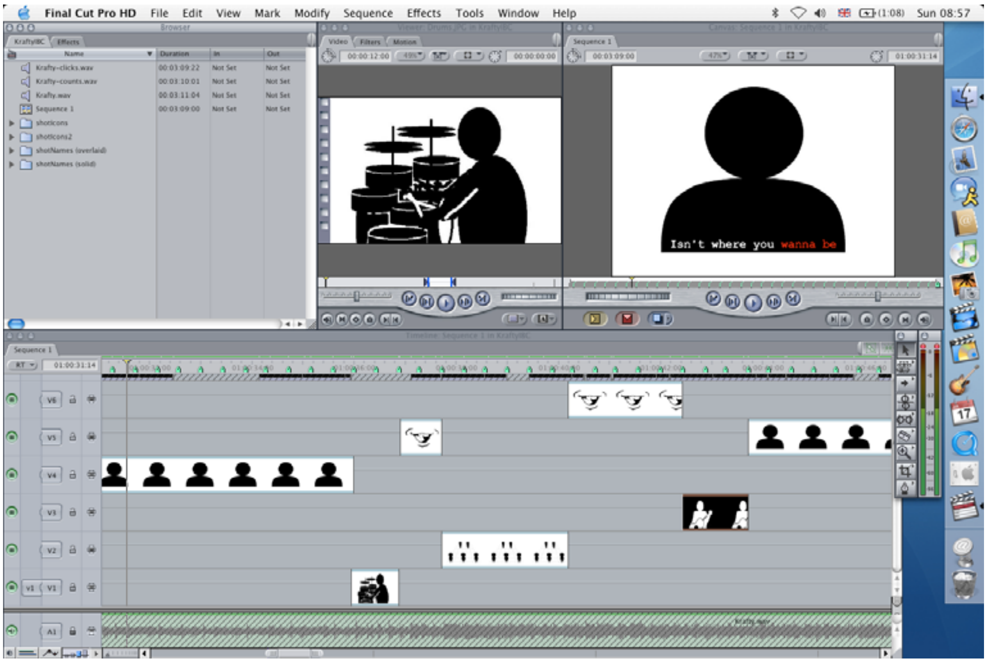


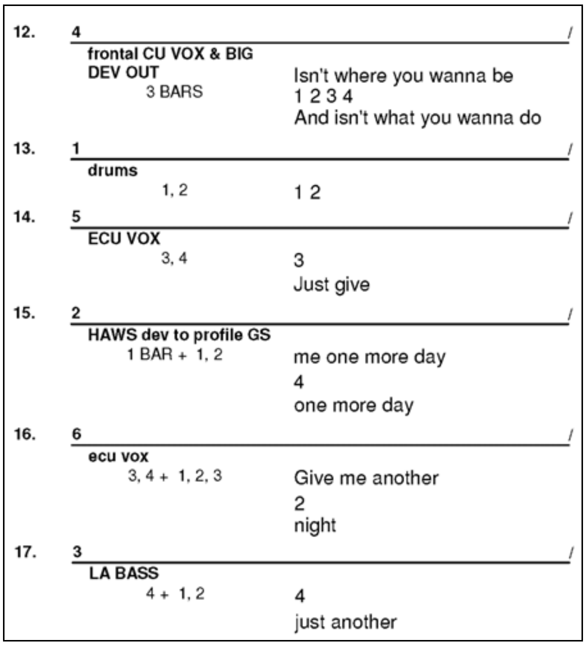
Figure 6 – The video timeline used as a shot-planning tool

the shot images can be keyed over the top, with the lyrics text track as the topmost layer. When the timeline is replayed, the dance rehearsal can be viewed with the shot details and lyrics superimposed and this helps considerably in visualising the final effect.

Creating the script

Once the Director has completed planning the shots, the three sequences of timings that are required for the final camera script are available: the beats (in milliseconds), the word timings (in milliseconds) and the shot timings (in frames). The final step is to merge these three to produce the script. As the script uses beats as its unit of time, all three are expressed in terms of beats first. The script generation algorithm then iterates through all the beats and checks whether a word is being sung during that beat, and whether a shot change occurs after that beat. To create the final script, we first take the beats and word timings and show the words where they are sung, otherwise we show the beat numbers. Then we bring in the shot timings. If a shot change occurs after the current beat, a shot separator showing the camera number and details is inserted into the script.

This final step of creating the script is automatic and currently produces the script in the BBC "ScriptWriter" format. Figure 7 shows part of a script that has been created in this way. Each camera shot has been added to the music breakdown and any camera directions, added by the Director in the timeline, carried through to appear beneath each cut line. In addition, the duration of each shot, in bars and beats, is also added automatically.



12. 4 /
frontal CU VOX & BIG
DEV OUT
3 BARS
Isn't where you wanna be
1 2 3 4
And isn't what you wanna do

13. 1 /
drums
1, 2
1 2

14. 5 /
ECU VOX
3, 4
3
Just give

15. 2 /
HAWS dev to profile GS
1 BAR + 1, 2
me one more day
4
one more day

16. 6 /
ecu vox
3, 4 + 1, 2, 3
Give me another
2
night

17. 3 /
LA BASS
4 + 1, 2
4
just another

Figure 7 – A script produced automatically through the new planning system.

After the planning stages are complete, the script that is produced is in the same format as one created manually and so the remainder of the production cycle is unchanged. The scripts are printed, camera cards produced and the production, when it reaches the studio, is no different from one planned in the traditional way.

PRACTICAL EXPERIENCE SO FAR

Working with the production team for the weekly pop music programme *Top of the Pops*, it soon became clear that this new approach offers them considerable scope for reducing the time spent on the repetitive parts of the planning process. The beat and bar analysis worked well and the users were quickly able to produce a music breakdown. The main bottleneck was the lyrics alignment for those cases where the speech recogniser initially had difficulties and re-voicing the lyrics was necessary to overcome this. We hope that various planned improvements to our software will reduce this problem considerably.

The programme Director was very enthusiastic about this new way of planning using the timeline editor. The visualisation of the camera shots helped to get a far better feel for how sequences of shots interact with each other. The instantaneous visualisation allowed shots to be refined during the planning process, something that would normally not be possible until the rehearsals in the studio. As a consequence, we expect the quality of the camera scripts produced this way to be higher, with fewer changes being necessary during rehearsals in the studio. However, if last-minute changes are required, the new system makes it easier to apply them without the need to re-process the entire music track.

Even at this very early stage, the extra effort needed to get used to the new way of working was offset by being able to generate the camera script with shot timings by simply clicking a button - it was with some amazement that they saw what would normally take several hours being reduced to one click of a button.

FUTURE WORK

Building on these results and with the feedback that we have had from colleagues in production areas, we intend to develop this work in three areas.

Firstly, we have focussed so far on pop music production but would like also to explore how these ideas can be applied to classical music programmes.

Secondly, the commercial video editing systems that we have explored are not necessarily optimised for planning in the way we would like to use them. We would therefore like to see how they might be extended to provide more convenient facilities for this type of application.

Finally, we have received several comments from production staff that this same approach could well offer benefits for other programme types, such as drama. Our future work will therefore also look at how TV production planning in general can benefit from these ideas.

REFERENCES

1. Scheirer, E.D., 1998, Tempo and beat analysis of acoustical musical signals, Journal of the Acoustical Society of America, Vol. 103, no. 1, pp. 588-601

2. Evans, M.J., 2003, Speech Recognition in Assisted and Live Subtitling for Television, BBC R&D White Paper WHP065, http://www.bbc.co.uk/rd/pubs/whp/whp065.shtml

ACKNOWLEDGEMENTS

We would like to thank our colleagues Jon Attard, Tony Gregory and Tony Grech for their helpful discussions, enthusiasm and encouragement in developing the ideas described in this paper.