IET Journals
The Institution of Engineering and Technology

Vol 5 | ISSN 2041-5923
2013

The Best of IET and IBC

INSIDE Papers and articles on electronic media technology from IBC2013 presented with selected papers from the IET's flagship publication *Electronics Letters.*

Contents

An electronic version of this publication and previous volumes can be found at www.theiet.org/ibc or by scanning the QR code

ISSN 2041-5923

Introduction

Welcome to *The Best of IET and IBC*, 2013. This is the fifth volume[1] of an annual joint publication between the International Broadcasting Convention and the Institution of Engineering and Technology.

The IET is a formal member of the IBC's partnership board but, beyond this, it has a long-standing and close relationship with the organisation, through which they together encourage and promote professional excellence in the field of media technology. Nowhere is this relationship more strongly reflected than in the pages of this publication, which celebrates the very best technical media papers from this year's IBC and the IET's flagship journal, *Electronics Letters*.

This year, our editorial takes a look at IBC's Future Zone – a real 'ideas-fest' where research organisations and collaborative consortia proudly show off some of their latest concepts in media technology. Few visitors to the Zone remain unmoved by the experience and talk among engineers at IBC is very often about the novel and exciting prototypes they've seen demonstrated there.

We then present eight papers chosen as the best contributions to IBC2013 by the IBC Technical Papers Committee in conjunction with the executive team of the IET Multimedia Communications Network. These include the overall winner of IBC's award for the Best Conference Paper 'New-generation Scalable Motion Processing from Mobile to 4k and Beyond' and papers representing the hot topics of 2013: HEVC (High Efficiency Video Coding), UHDTV (Ultra High Definition Television), (HDR) High Dynamic Range video, the IP studio, access services and new efficient ways of using terrestrial spectrum.

We include an interview with Jorge Rodríguez, winner of IBC2013's best young professional contribution as lead author of 'Challenges and Opportunities for Successful UHDTV Broadcasting', which is also published in this volume.

From *Electronics Letters* this year we have three papers chosen from those published since IBC 2012, one of which includes a supplementary article, a feature introduced in *Electronics Letters* to provide readers with more background and behind-the-scenes information on some of the research published in each issue.

Just when we thought that broadcasting technologies had reached a pinnacle in the efficiency of their delivery, it is amazing to see how the world's media specialists have managed to find and exploit further hidden pockets of redundant information and spectral capacity. All this will have a very profound and positive effect on the economics and speed with which we can expect to be enjoying novel and exciting services, both in our homes and while on the move.

I hope that you enjoy reading this collection of the best papers as much as I and my committee of specialists and peer reviewers. We would like to convey our thanks to everyone involved in the creation of this year's volume and extend our best wishes for a successful and stimulating IBC 2013.

Dr Nicolas Lodge
Chairman IBC Technical Papers Committee
&
The IET Multimedia Communications Network Executive Team

[1]For previous volumes see www.theiet.org/ibc.

Who we are

IBC

IBC is committed to staging the world's best event for professionals involved in content creation, management and delivery for multimedia and entertainment services. IBC's key values are quality, efficiency, innovation, and respect for the industry it serves. IBC brings the industry together in a professional and supportive environment to learn, discuss and promote current and future developments that are shaping the media world through a highly respected peer-reviewed conference, a comprehensive exhibition, plus demonstrations of cutting edge and disruptive technologies. In particular, the IBC conference offers delegates an exciting range of events and networking opportunities, to stimulate new business and momentum in our industry. The IBC conference committee continues to craft an engaging programme in response to a strong message from the industry that this is an exciting period for revolutionary technologies and evolving business models.

The IET

The IET is one of the world's leading professional societies for the engineering and technology community, with more than 150,000 members in 127 countries and offices in Europe, North America and Asia-Pacific. It is also a publisher whose portfolio includes a suite of 27 internationally renowned peer-reviewed journals covering the entire spectrum of electronic and electrical engineering and technology. Many of the innovative products that find their way into the exhibition halls of IBC will have originated from research published in IET titles, with more than a third of the IET's journals covering topics relevant to the IBC community (e.g. *IET: Image Processing*; *Computer Vision*; *Communications*; *Information Security*; *Microwave Antennas & Propagation*; *Optoelectronics*, *Circuits & Systems* and *Signal Processing*). The IET Letters contained in this publication come from the IET's flagship journal, *Electronics Letters*, which embraces all aspects of electronic engineering and technology. *Electronics Letters* has a unique nature, combining a wide interdisciplinary readership with a short paper format and very rapid publication, produced fortnightly in print and online. Many authors choose to publish their preliminary results in *Electronics Letters* even before presenting their results at conference, because of the journal's reputation for quality and speed. In January 2010 *Electronics Letters* was given a fresh new look, bringing its readers even more information about the research through a colour news section that includes author interviews and feature articles expanding on selected work from each issue.

Working closely with the IET Journals team are the IET Communities team. The communities exist to act as a natural home for people who share a common interest in a topic area (regardless of geography); foster a community feeling of belonging and support dialogue between registrants, the IET and each other. Assisting each community is an executive team, made up of willing volunteers from that community who bring together their unique experience and expertise for the benefit of the group. Members of the Multimedia Communications Community executive team play an essential role in the creation of this publication in reviewing, suggesting and helping to select content. They contribute their industry perspectives and understanding to ensure a relevant and insightful publication for the broad community represented at IBC, showing the key part volunteers have to play in developing the reach and influence of the IET in its aim to share and advance knowledge throughout the global science, engineering and technology community.

ISSN 2041-5923

Editorial

The Future Zone at IBC – where tomorrow comes alive

The Future Zone is an exciting concept at IBC that has grown both in size and visitor interest over the years. It provides a tantalising glimpse into the future technologies in the broadcasting and electronic media industries. The Future Zone is located outside Hall 8, next to the Park Foyer, and occupies this year some 150 square metres of floorspace.

The layout of the Zone allows for a number of stands to be located around the perimeter, where companies, consortia and universities can demonstrate the latest conceptions and prototypes from their labs. The only rule of engagement is that products that already exist on the market cannot be shown; and this creates an ethos in the Future Zone that is similar to international motorshows with their 'concept car' models – and this is the opportunity to talk to the engineers and developers themselves.

This year's exhibitors with demonstrations in the Future Zone include:

- international collaborative research consortia like SCENE, TOSCA-MP, Vista-TV and HBBNext with their ideas for next generation media production and smart TV systems;

- the world's top research laboratories, like BBC, ETRI, NHK, showing their latest wares relating to future beyond-HD media content and services;

- companies like ACB and Intel demonstrating new immersive display concepts and consumer interaction techniques;

- universities, represented by Braunschweig with some disruptive technology proposals for broadcast networks, and Warwick spin-out goHDR with its high dynamic range video expertise;

- organisations working in 3D audio and producing new dimensions in audio quality, which can be heard in Pinguin's and BiLi's stands.

Within the core of the Future Zone, there is also a strong cohort of over a dozen exciting poster displays, exhibiting innovations from around the world, which are still at the ideas stage. The posters area enables IBC Conference to create a setting where a variety of different technical subjects can be shown all together, allowing topics to be aired that might not otherwise be able to be accommodated within the Technology Stream agenda. This year's posters area has a collection of mind-bending concepts and ideas, and provides the ideal opportunity again for visitors to discuss questions directly with the inventors themselves.

The Future Zone is open throughout the IBC main Exhibition times, but the posters are manned only during Friday and Saturday. A highlight event is scheduled in the Future Zone between 16:00 and 18:00 on Friday the 13th September; this is a special IET-sponsored Champagne Reception, launching the latest edition of the popular technical journal "Best of IET & IBC" and reviewing all tomorrow's technologies on show in the Zone.

Professor David Crawford
IBC Conference Producer,
Technology Stream

Selected content from IBC2013. Winner of the Best Paper award

ISSN 2041-5923

New-generation scalable motion processing from mobile to 4K and beyond

Mike Knee

Snell Ltd., UK

Abstract: Today's broadcast video content is being viewed on the widest range of display devices ever known, from small phone screens and legacy standard definition TV sets to enormous 4K and 8K UHDTV displays. The growth in size and resolution is happening alongside many other improvements, in grey-scale resolution, colorimetry, 3D and, especially, higher frame rates. These developments mean that the requirements for very high quality, artefact-free conversion in resolution and frame rate have become more important than ever. The challenge is given a further dimension by the wider range of content that can appear on large screens, from upconverted archive footage to the much more detailed, wider window on the world made possible by the new large formats.

This paper presents cutting-edge algorithms for motion compensated processing to meet these challenges in both live TV and file-based operation. One size no longer fits all, so this paper also discusses how to achieve a balance across the range of processing complexity and performance, showing how the trade-offs can be managed gracefully and optimally.

1 Introduction

How will you watch your next TV programme? Will it be on a small phone screen, a head-up display, a tablet, an old CRT TV, a PC monitor, a modern HDTV display, a projector, a 4K or an 8K UHDTV display? And where will the content have come from? A mobile phone video, an old SD TV drama, an HDTV production studio, a digital film master or a 4K or 8K camera? We rightly expect seamless transfer of content from all those sources to all those destinations, and for differences in colorimetry, dynamic range, resolution, interlace, aspect ratio and frame rate to be dealt with efficiently, without loss of image quality or visual impact.

In previous IBC papers, we have looked at HDTV standards conversion [1], interlaced and progressive signals [2] and novel ways of processing material for smaller and different-shaped displays [3]. Those technologies and algorithms continue to be relevant. However, in recent years the question of field or frame rate has become increasingly important, as interest has grown in conversion not only between the standard field rates of 50Hz and 59.94Hz, but also from and to 24Hz and newer film frame rates such as 48Hz, and higher frame rates in cameras and displays such as 100Hz, 120Hz, 300Hz and beyond. One particular example of interest is conversion from 24Hz film to 50Hz and 59.94Hz frame rates, in a world that is becoming increasingly intolerant to the motion judder resulting from conventional 2:2 and 3:2 pulldown methods of conversion.

Motion compensated processing has long been considered essential for high-quality frame-rate conversion. However, the massive increases in screen size, resolution and display brightness have all put pressure on the previous generation of motion compensated algorithms. A step change in motion compensation technology is required to meet these new demands. At the same time, cost pressures on programme production and distribution in multiple formats are bringing a requirement for greater flexibility in allocation of resources to tasks such as conversion in both live and file-based applications.

This paper presents a new generation of algorithms for motion compensated processing. First, we look at a particular problem that has emerged as the range of source and display resolutions increases, to describe which we have adopted the term 'wow factor', and which the new algorithms are particularly suited to address. We then look at developments in the two main components of motion compensated processing: motion estimation and picture building. Finally, we introduce the concept of a single ‘knob’ which can be used to control the trade-off between processing speed and conversion quality, and discuss how to perform scalable load balancing using available processing resources across varied input picture content.

2 Window on the world

2.1 Range of resolutions

The range of source and display resolutions we might encounter is now very wide. A small mobile phone might have as few as 0.1 Mpixels, while with 8K UHDTV we have 32 Mpixels, a ratio of 320:1. At any display resolution, it is important to ask ourselves where the source has come from, in particular what resolution it was captured at, and also what production techniques were used. We shall now look at these questions with particular reference to pictures that are displayed at high resolution, taking 8K as an example.

2.2 Low-resolution source

A source at a low resolution, for example standard definition, will normally be upconverted if it is to be displayed at high resolution. Typical SD camera techniques involve zooming-in quite close to the subject. Any motion in the source will be, in pixel terms, faster in proportion to the degree of upconversion in each dimension, and the large picture will cover a fairly small viewing angle in the original scene and will be relatively soft.

2.3 High-resolution source

If the source is at high resolution, it will be displayed unchanged on the high-resolution display, and the characteristics of what is displayed depend on the production technique. If, on the one hand, the camera is used as if it were a low-resolution camera, the picture will have the same characteristics as one from the low-resolution source. On the other hand, the viewing angle of the camera could be widened so that the high display resolution is fully exploited, in which case the picture will typically contain more detail, smaller objects and lower motion speeds.

2.4 The ‘wow factor’

We propose a simple rule of thumb for expressing the different possible picture characteristics seen in high-resolution displays. The ‘**wow factor**’ (**w**indow **o**n the **w**orld) indicates the degree to which increased display resolution is exploited to give the viewer a wider view of the scene. An example showing the relationship between display format, upconversion ratio and wow factor is shown in Fig. 1. The diagram shows that, as the display format grows, the range of possible wow factors increases. Table 1 summarises qualitatively the effect of the wow factor on parameters relating to motion compensated processing.

2.5 Motion compensation

This analysis has unveiled a problem that occurs when it comes to scaling up a motion compensated processing algorithm for larger display formats. If the wow factor remains low, the processing will have to cope with fast motion of blurred objects. If it is increased, the processing will have to cope with small, detailed objects. Of course, in reality we have to cope with the full range of wow factors, which doubles for every doubling of the display resolution. This means that scalability of motion compensated processing becomes a multi-dimensional affair and will not be handled satisfactorily by any single scaling up of a standard definition (SD) or high definition (HD) system.

We now discuss the effect of these observations on motion estimation (the analysis part of motion compensated processing) and on picture building (the synthesis part) in turn.

3 New-generation motion estimation

So how do we design an improved and scalable motion estimator? Here we introduce some of the new approaches we have made towards a fully scalable algorithm.

After several decades of research, methods of motion estimation [4] still largely fall into the categories of block matching, gradient or optical flow methods [5], frequency-domain methods such as phase correlation [6] and feature-based methods. The new suite of algorithms presented here, code-named ‘Mensa’, makes extensive use of the first three categories, while work is proceeding on introducing the fourth category into the mix.

3.1 Multi-scale candidate vector generation

Our existing motion estimation technology makes use of phase correlation to analyse a scene and to generate candidate motion vectors for subsequent assignment to individual pixels. The phase correlation is based on large blocks, whose size is a trade-off between motion range and ability to handle small or detailed objects. We have seen that both are required, so Mensa uses multiple block sizes in parallel to generate candidate motion vectors of both kinds.

3.2 Gradient-based refinement

One disadvantage of phase correlation is its fundamental inability to handle smooth variations of motion within objects, such as zooms, rotations and perspective in a receding landscape. Where the ‘wow factor’ is low, this does not pose too great a problem, because the degree of

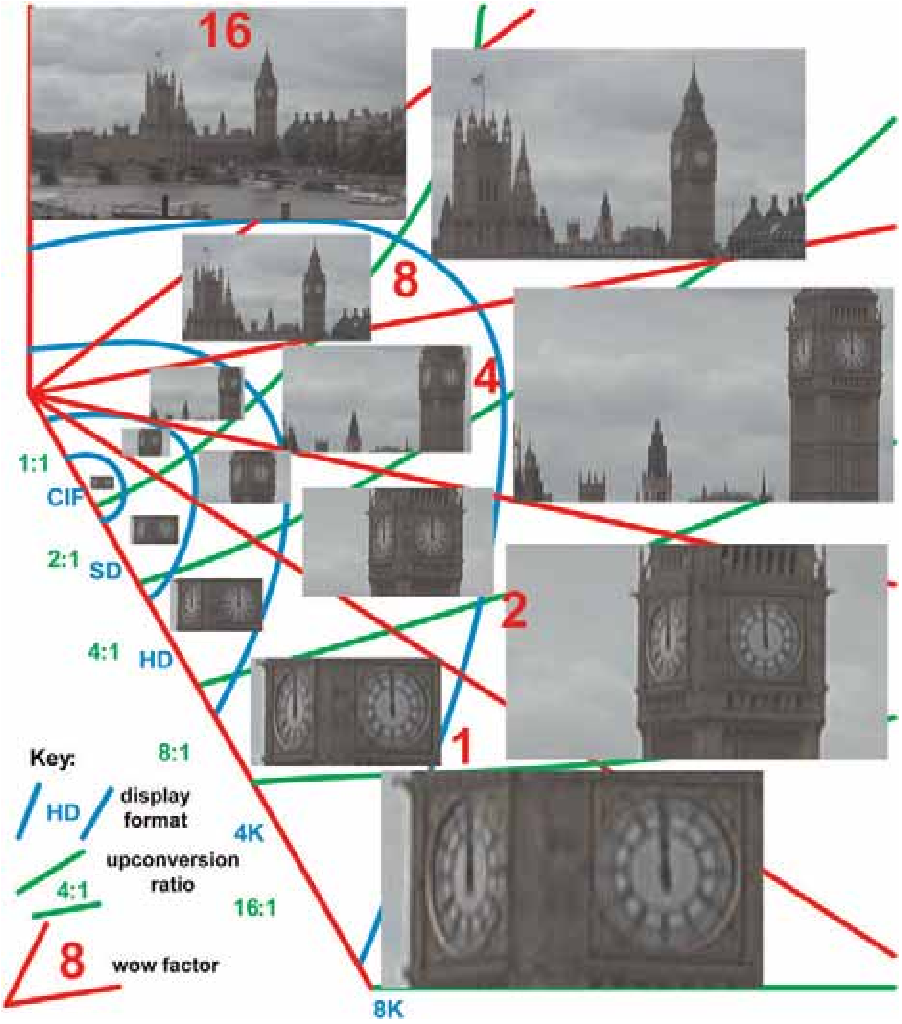


Figure 1 *Wow factors*

motion variation is also low. But for high wow factors these variations can become quite large. We have solved this problem by allowing the candidate motion vectors to vary slowly from pixel to pixel, using gradient based techniques to refine the vectors from initial constant values.

One of the drawbacks of gradient-based vector refinement is that it fails at motion boundaries. We overcome this problem by using overlapping blocks, with a weighting system to encourage refinement to work hardest in areas where the vector field is already performing well. This has the effect that one overlapping block covering a motion boundary will refine a vector field suitable for one side of the boundary but not necessarily the other, while another overlapping block will refine a field for the other side.

Table 1 Effects of 'wow factor'

Wow factor	Low	High
Sharpness	Low	High
Object size	Large	Small
Motion speed	Fast	Slow
Motion variation	Narrow	Wide

3.3 Vector assignment

The final step in motion estimation is to assign a motion vector to every pixel from the set of refined candidates. The classic way to assign motion vectors is to calculate an error surface for each candidate, usually a displaced frame difference (DFD). This surface is then filtered spatially so that the error generated by a candidate for a pixel contains information about the neighbourhood of the pixel. It is difficult to choose the right size for this filter: too small, and the vector field is noisy; too large, and small objects can be missed and behaviour at motion boundaries is poor. For Mensa, we have developed a nonlinear DFD filter based on splitting the neighbourhood into octants, as shown in Fig. 2, and applying a minimax approach that allows a motion boundary to pass through the neighbourhood while retaining the stability of a large filter.

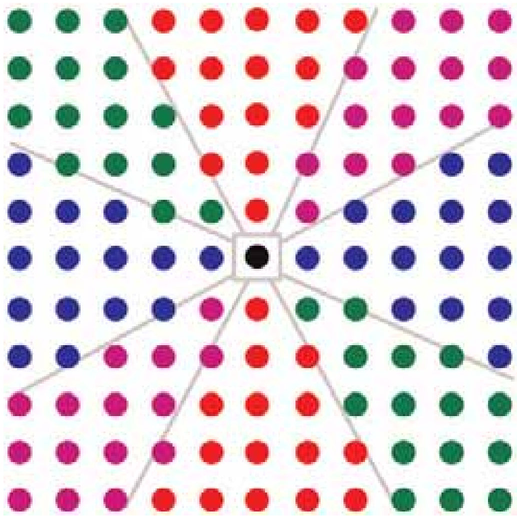


Figure 2 *Octant filter*

4 New-generation picture building

The second main component of a motion compensated processing system is a rendering engine or 'picture builder' which takes the input pictures and associated motion vector fields and uses them to build an output picture at a desired time instant. Because motion vectors are associated with input pixels and not output pixels, a projection operation is required in which input pixels are written into the output picture at locations determined by the motion vectors and the desired temporal phase. Such projection requires mechanisms for handling occlusions, multiple 'hits' from different input pixels, 'holes' where an output location is not written, and sub-pixel interpolation. It is at this stage that any problems resulting from inaccurate motion vectors, transparency, very complex motion and other transformations in the picture, may appear as annoying artefacts.

4.1 Wavelet picture building

It is possible to manage the appearance of these artefacts and to reduce their overall visibility by employing a 'wavelet picture builder' in which the output picture is built up in sub-bands with suitably scaled and downconverted motion vectors at each stage. A simplified example of one layer of the Mensa wavelet picture builder is shown in Fig. 3. A feature of this approach is that holes in the projections are automatically filled from coarser layers.

4.2 Temporal phase control

There are some kinds of picture material that will defeat even the most reliable motion estimators and the most benign picture builders. It is prudent to have recourse to some kind of 'fallback mode' which is applied when such picture material is encountered. Crucial to the usefulness of a fallback mode is a reliable metric that will determine when and to what extent it should be applied. Our metric is based on the assignment errors, and the principle of our fallback mode is to build pictures that are closer in time to the input pictures.

5 The Mensa Knob

In this section, we turn to some work on optimising the cost/performance trade-off of a complex machine such as the standards converter described above. When a standards converter is implemented in hardware, the full resources of

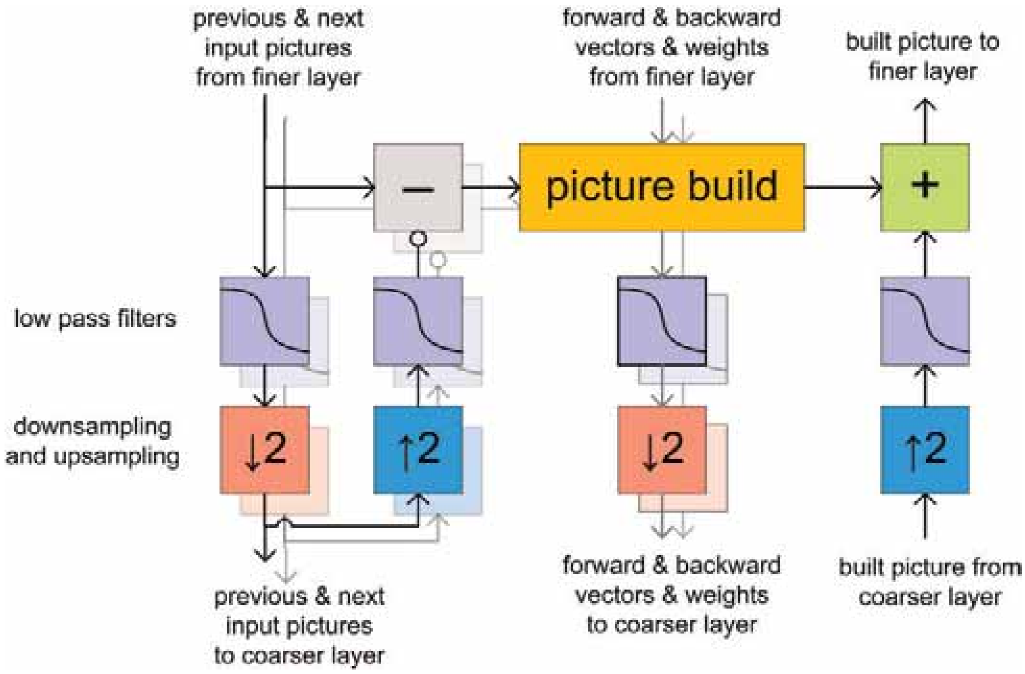


Figure 3 *One layer of a wavelet picture builder*

a complex algorithm can be applied without inefficiency (except possibly in electrical power) to both demanding and easy picture material. But in an implementation based on software, processing time and the number of processors required are directly measurable as a processing cost, and it becomes beneficial to tailor the processing to the content.

5.1 Streaming and file-based processing

Some applications of video processing are designed for real-time streaming, usually with a limit to the permitted latency. Others are file-based and may work faster or slower than real-time. In both cases there is scope for optimisation of the performance/cost trade-off, though the possibilities are greater in the case of file-based processing. For a given set of content, there may on the one hand be a limit to the time and processing resources available, and the goal is to maximise the quality of the output pictures. On the other hand, there may be a required minimum quality level, and the goal is to minimise the processing time or number of processors used in order to save time and money. But even for live streaming, it may be possible to concentrate resources on locally more demanding parts of a video stream.

5.2 The efficiency cloud

A conversion algorithm such as Mensa is controlled by a multitude of parameters. Some of them, such as thresholds or gain factors, will typically only affect performance and have no impact on processing time. These can generally be optimised in a straightforward manner, given a suitable performance metric, though it may be worth repeating the optimisation process for different genres of input material, for example sport or news. Other parameters, such as numbers of candidate motion vectors or of vector refinement iterations, will generally affect both the performance and the processing time. The interactions between these parameters can be bewilderingly complicated, making it very difficult to control the performance/cost trade-off. Fig. 4 shows the results of processing a test sequence with hundreds of combinations of control parameters. The *x*-axis represents the processing time (the scale is arbitrary) and the *y*-axis represents a performance error measure, in this case the root-mean square (RMS) error between the output sequence and a known 'ground truth' sequence. Note the false origin on the *y*-axis, highlighting the fact that small (though visible) performance improvements are generally only obtained at the cost of substantial increases in processing time. A few points extend above and to the right of the cloud shown.

It would be highly desirable to reduce the set of adjustable parameters to just one: a single controller or 'knob' which could be adjusted between relatively high errors but low processing cost, and low errors but high processing cost. This could be achieved by selecting a subset of points in the cloud that span the range of performance and processing time but which are in some sense optimal. Looking at Fig. 4, it becomes clear that some parameter selections are less efficient than others. For example, point A has both a higher processing time and RMS error than point 5, so within the assumptions we have made, point A would be of no use in a knob. Suitable points would be those that are on the approximately hyperbolic envelope of the left and bottom of the cloud.

5.3 A knob that goes to 11

Fig. 4 shows a labelled subset of points that follow the envelope and which would therefore make good

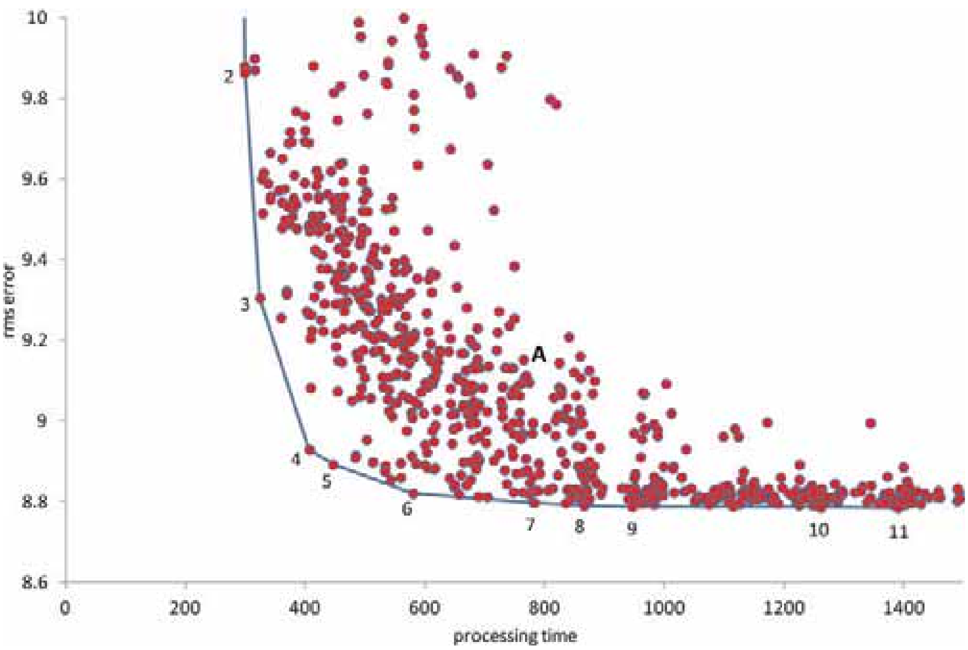


Figure 4 *The efficiency cloud and the Mensa Knob*

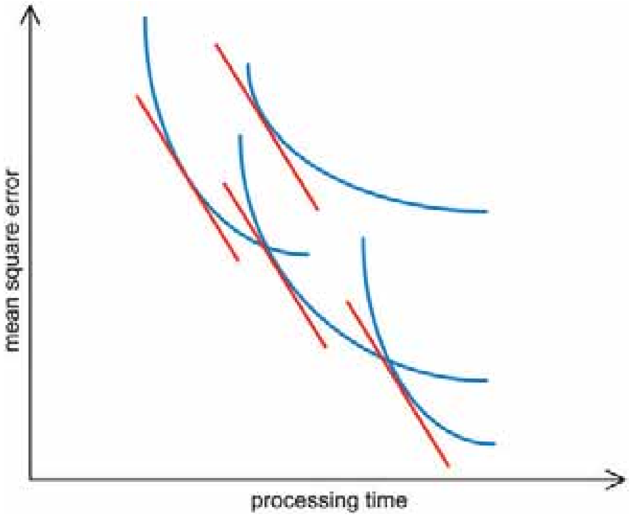


Figure 5 *Load balancing*

candidates for a performance knob. Point 0, which is non-motion-compensated conversion, and point 1, a very simple motion compensated algorithm, fall well above the top of the plot. The fact that the knob settings extend to 11 is a serendipity, echoing the scene in the cult 1984 film 'This is Spinal Tap' in which a joke is made about amplifier knobs that 'go to 11' rather than the standard 10.

Each knob setting maps to a selection of parameter choices, and it is now possible to make adjustments between high performance and high speed, knowing that each setting is performing at optimum efficiency.

5.4 Scalable load balancing

The above analysis is based on an ensemble of test material of varying degrees of difficulty. In practice, the performance of a particular knob setting will depend on the source material. Whether our goal is to minimise overall error given a processing time limit, or to minimise processing time given a maximum acceptable error, we need an algorithm that links some measure of source difficulty to the knob setting.

If we repeat the analysis for different sources, we would obtain a set of different 'knob' curves, as illustrated in blue in Fig. 5. Note that the y-axis now represents mean 'square' error, so that errors can be added across all the sources.

Suppose for each source i the mean square error e is linked to processing time t by a function

$$e = f_t(t)$$

If each source has M_i frames, then the total error is

$$E = \sum_{i=1}^{N} M_i f_i(t_i)$$

and we wish to choose t_i, the processing time per frame for each source, to minimise E subject to a total processing time constraint:

$$T = \sum_{i=1}^{N} M_i t_i$$

Using the method of Lagrange multipliers, the equations to solve are:

$$\frac{\partial E}{\partial t_i} + \lambda \frac{\partial T}{\partial t_i} = M_i \left(\frac{\partial f_i}{\partial t_i} + \lambda \right) = 0, \quad \text{so} \quad \frac{\partial f_i}{\partial t_i} = -\lambda$$

which just means that we have to choose points on each function where all the gradients are the same, as shown by the red lines, and the choice of gradient will be such as to meet the total processing time limit. Similar reasoning would apply to meet an error constraint.

The remaining problem is to find out, given real picture material, which curve is appropriate for each source segment. We no longer have ground truth, and we certainly cannot afford to try out different knob settings, so we have to gather evidence by taking measurements on the source pictures. For example, we can calculate the average frame-to-frame difference of each segment. It turns out that there is a reasonable correlation between such a simple measure and the 'knob function'. This allows us to choose an appropriate knob setting for each segment in order to optimise the overall cost/performance trade-off.

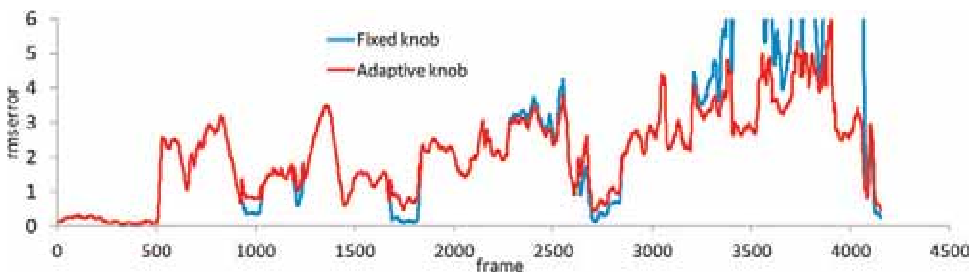


Figure 6 *Load balancing example*

Fig. 6 shows a comparison between this load-balancing approach and a fixed knob setting with the same overall processing costs. The graphs show the RMS error for a three-minute section of a 1960s spaghetti western film when converted from 24 to 60Hz using knob settings at the lower end of the processing quality scale. For the purposes of this illustration, the 'ground truth' is taken to be the output of knob setting 11, a technique which turns out to be remarkably useful when evaluating the lower-quality settings.

In this example, the error for some of the 'easy' segments has been allowed to increase, freeing up processing time to improve the performance of the most difficult segments.

6 Conclusions

In this paper, we have introduced a new generation of motion compensated processing algorithms suitable for the very wide range of source and display resolutions now encountered, and have described how they can be controlled in such a way as to optimise the performance/cost trade-off in both streaming and file-based processing.

7 Acknowledgments

The author would like to thank the Directors of Snell Ltd. for their permission to publish this paper, and his Snell Technology Development Algorithms team colleagues for their valuable contributions, suggestions and support. Intellectual property disclosed in this paper is the subject of patent applications and granted patents in the UK and elsewhere.

8 References

[1] KNEE M.J.: 'International HDTV context exchange'. Proc. IBC, 2006

[2] KNEE M.J.: 'Progressive HD video in the multiscreen world'. Proc. IBC, 2010

[3] KNEE M.J., PIRODDI R.: 'Aspect processing: the shape of things to come'. Proc. IBC, 2008

[4] DUFAUX F., MOSCHENI F.: 'Motion estimation for digital TV: a review and a new contribution'. Proc. IEEE, June 1995, vol. 83, no 6

[5] BLACK M.J., ANANDAN P.: 'A framework for the robust estimation of optical flow'. IEEE Proc. Fourth International Conference on Computer Vision, 1993

[6] ARGYRIOU V., VLACHOS T.: 'A study of sub-pixel motion estimation using phase correlation' (Centre for Vision, Speech and Signal Processing, University of Surrey, 2006)

Selected content from IBC2013

ISSN 2041-5923

The IP Studio

P.J. Brightwell   J.D. Rosser   R.N.J. Wadge   P.N. Tudor

BBC, UK

Abstract: As the performance and cost-effectiveness of packet-switched networks continue to increase, broadcasters can benefit from handling real-time production-quality video and audio as first-class citizens on IP infrastructure. When combined with capturing richer production datasets, this provides new opportunities to produce innovative, and more personalised, products. BBC Research & Development is developing a framework to investigate going beyond traditional technologies such as SDI towards an IP-based infrastructure, using widely adopted networking protocols and techniques. The framework treats items of content, such as video frames, as well as data events generated in a production, as uniquely identified and timestamped objects. For live working, these are multicast at high quality locally and published as web feeds for remote access; they are also stored for immediate use by downstream processes and devices, without the overheads associated with file-based transfers.

1 Introduction

The broadcast business has a track record of being changed by IT, and benefiting from the economies of scale in that industry. Home and mobile broadband have revolutionised consumers' access to content, and modern content distribution networks have transformed how broadcasters can deliver it. Production has also changed, and it is now usual to record onto files rather than tapes. However, in many cases the activities involved in file-based production tend to be based around processes acting on entire recorded clips, often causing unwanted delays in a team's operations. This is particularly true for live and near-live productions.

Traditional interconnect technologies such as SDI, AES-3 and SMPTE timecode have proved to be effective at their intended use, but their 'disconnect' with more mainstream network technologies imposes overheads. In the past, the prohibitive cost of multi-gigabit network infrastructure meant there was no practical alternative. Typical modern production facilities now include extensive IP networking, and the possibility of extending this to handle real-time production quality video and audio directly is enticing. This allows the network itself to handle routing of signals, using proven approaches to multicast distribution and security where required.

The potential benefits extend beyond savings from reduced specialist infrastructure: a converged platform could allow teams earlier access to their content, and more flexibility in how they can use it. Productions would be able to capture richer production datasets, aiding creation of more personalised content and new types of product. And by exploiting techniques for dynamic provision and control of networked computing resources, the broadcast industry could also benefit in terms of how it procures its facilities.

A major aim of BBC Research & Development's current work is to examine how the meaning of the term 'broadcasting' will change in the age of the Internet [1]. As part of this, the IP Studio project [2] is developing a framework consisting of a content model, transport mechanisms, system architecture and implementations to investigate the benefits and challenges of adopting an all-IP approach to production infrastructure, with a particular emphasis on live and near-live working. This paper presents the technical approach taken and demonstrates examples of how future production users and audiences could benefit from the possibilities offered.

2 Content model

2.1 Sources, flows and grains

At the heart of the IP Studio framework is the idea that everything of interest generated in a production should be captured and made available for immediate or later access.

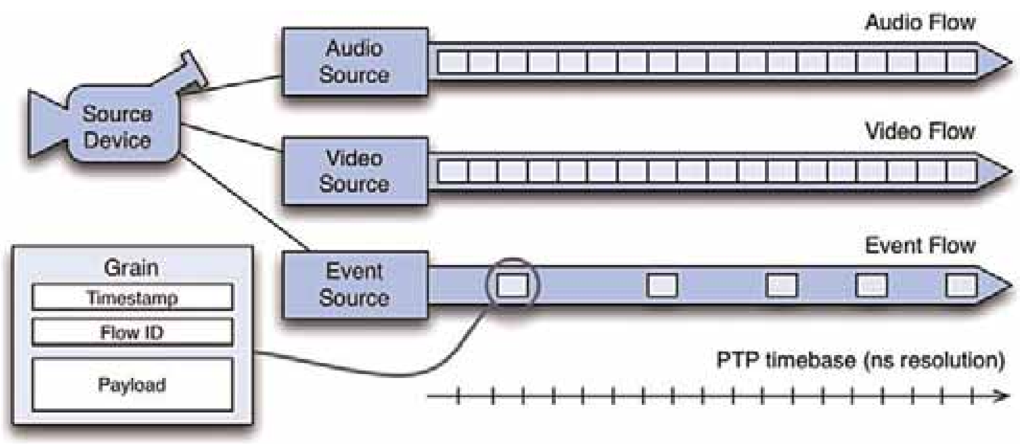


Figure 1 *Sources, flows and grains*

This goes beyond just audio and video, and typical 'clip-oriented' metadata, to include **events**: time-related data objects that contain information such as: camera settings, actor positions, logging entries and AV quality analysis data. In the project's content model, events, and frames or sections of video and audio, are all treated as individually identifiable **grains**. Grains are time-stamped objects within **flows** of time-sequential information coming from **sources** (see Fig. 1). An application can access grains in real time as they are created, or at a later time for retrieval from a particular timestamp.

2.2 Relationships

Each source and flow is uniquely identified (with a UUID), allowing a system to determine the origin of any grain. Information about the relationships between sources, and with other things such as cameras and people, is also recorded. These data allow complex queries to be constructed. As an example, a user might want to view what was recorded from each camera when a particular line of dialogue was spoken. This could be achieved by: querying a database to find a particular (speech recognition) event grain; finding its corresponding source; finding related sources; and finding video grains with matching timestamps.

2.3 Synchronisation

Synchronisation between grains is achieved using a common time reference. Typically computer clocks are synchronised using Network Time Protocol (NTP), but as this does not provide the level of accuracy that will be needed for many professional applications, the project is investigating the use of Precision Time Protocol (PTP) [3]. Accordingly grains have nanosecond-precision timestamps. They also optionally can include SMPTE timecodes for use within existing infrastructures.

2.4 Addressability

Because sources, flows and other objects are uniquely identified, they can be addressed using URIs, for example:

http://www.example.com/source/103158A7-D6D1-4B77-BE56-A854CE6B2770

This approach can be extended to, for instance, allow a client to specify a URI that asks for content in a particular format, or play a stored item from a particular time. This opens the possibility of straightforward integration with web-based systems.

3 Processing

3.1 Pipelines and nodes

Flows of grains can be processed in several ways. For example, a video flow may be encoded into different formats, multiple flows may be composited and audio or video flows may be analysed to produce event flows. Such processing operations are combined into **pipelines** that run on computing instances, or **nodes**. Pipelines start and end by receiving and sending external flows from and to a network port, a pre-existing interconnect such as SDI or local storage on the node itself. Fig. 2 shows a simple example of a node and pipeline.

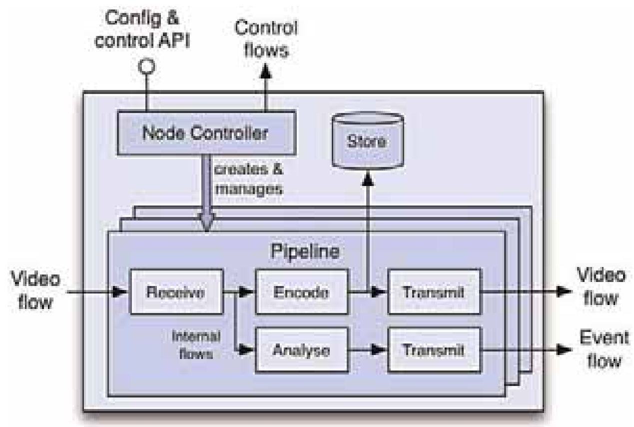


Figure 2 *Processing node*

3.2 Control

Each node provides a web service API to support remote configuration and control of the pipelines and their settings. The web service adapts automatically to reflect the pipelines running on the node. The API follows RESTful principles, responding to HTTP requests directly to resource URLs [4]. Clients can introspect the node to determine the structure of the pipelines and available parameters (e.g. encoder bit-rate), and dynamically build their user interfaces. Additionally, the node acts as a source of status events, allowing information about configuration changes to be propagated in real time around the studio, and stored for later use.

4 Streaming

Although SMPTE has defined a number of standards for wrapping video into IP, these are aimed at solving particular interconnection cases, such as the tunnelling of SDI through IP [5], rather than general transport of flows.

For low-latency applications within the IP Studio framework, for example monitoring, video and audio flows are streamed in elemental form (not multiplexed) using RTP [6]. Because there may be several devices receiving the same stream, the framework uses multicast extensively. Source-specific multicast [7] is preferred, to aid scalability and security.

A number of header extensions have been specified to carry the grain's identification and PTP time-stamp information. This provides the possibility of using RTP streams when a high accuracy of synchronisation is required. For such cases an accompanying RTCP channel is not used.

For other applications, for example when a reliable transport is essential, or when traversing firewalls, RTP may be less appropriate; so the IP Studio framework allows different transport mechanisms to be adopted. Of particular interest are MPEG-DASH [8], which is becoming increasingly relevant to content distribution, and HTTP Live Streaming [9], which is often used for streaming to portable devices. Note that both HLS and DASH use chunked HTTP transfers, making them unsuitable for low-latency applications. Similarly for events, WebSockets [10] provide a reliable transport to and from modern web browsers.

5 Software implementation

To demonstrate the feasibility of the techniques discussed here, the project has developed implementations of several aspects of the framework, and to potentially provide reference implementations for future standardisation activity. Where possible, open source components have been used; although proprietary modules such as video codecs can also be integrated.

A node's processing, local storage and control elements are implemented as compiled applications running as individual processes within a POSIX environment such as Linux. A node can be deployed on a standalone computer, or within a virtualised cluster.

A plug-in framework allows new types of processing elements to be incorporated in a straightforward way. Shared memory segments join processors together within pipelines, while the node controller uses a messaging mechanism [11] to control the processors, and lightweight HTTP and WebSocket servers for external configuration and control.

A 'polyglot' approach has been taken to storage [12]. A custom content store has been developed to provide rapid access to grains based on their flow identifier and timestamp, and different types of database are being assessed for their performance and scalability for event-based queries.

RTP streaming is provided via an open source library [13] that has been extended to incorporate the extensions outlined above. The project found that it was necessary to apply a traffic shaping algorithm to the egress of packets on sending nodes to prevent bursty RTP traffic from causing temporary packet loss.

As cameras and professional monitors do not yet come with streaming IP connectivity, a number of SDI and HDMI interface cards are supported. These are installed in dedicated servers running pipelines to produce AVC-I 100 [14] streams from 10-bit SDI signals. The project has experimented with other video formats, including MJPEG thumbnails and H.264 proxies. To date audio has been transported as linear PCM, with no limit on the number of channels; other audio formats will be supported in the future.

AVC-I encoded RTP streams to the project's specification can also be taken from a dedicated camera back node developed by BBC R&D's Stagebox project [15].

The IP Studio project has developed a number of native and web-based user tools for monitoring, logging and control. The example in Fig. 3 shows a multi-input viewer in which a tablet is used to control a receiver and decoder pipeline.

6 Example applications

The framework enables many production and audience use cases, some of which are discussed here.

6.1 Flexible viewing

Multiple video flows from the same source device can be created on demand for a range of purposes from viewing live thumbnails on a tablet up to critical monitoring (see

Figure 3 *Multiviewer and tablet controller*

Fig. 4). The pipeline parameters are set automatically based on the type of device and network connection.

6.2 Large-scale quality analysis

In a large facility, looking and listening for potential technical issues across many sources can be aided through analysis pipelines that generate events warning of problems. These can be aggregated and filtered to present a helpful summary for the operator, who can then review the suspect flow in more detail.

6.3 New forms of content

Camera positioning and pose information can be represented as event flows and used in conjunction with 3D models to produce synthetic renders on a variety of production and audience platforms.

6.4 Data-driven second screen

Face recognition and GPS location events can be captured during production and presented to users of second screen devices. The events are automatically linked and synchronised to the available content. This opens new opportunities for personalisation, such as allowing a user to follow a particular sport or person at a large event.

6.5 Social media events

Many teams already filter Twitter and Facebook feeds for comments about their programmes. IP Studio offers the possibility to connect this audience feedback directly into the live production environment.

7 Scaling up and out

7.1 Discovery

Once a facility grows to more than a handful of sources and nodes, configuration and management become difficult without help from the infrastructure. A key aspect is the automatic discovery when they are plugged in or provisioned. In the IP Studio framework, nodes send multicast DNS Service Discovery [16] advertisements about themselves, and about the sources they provide, out to the local network. However, it is often inappropriate for clients to deal directly with DNS messages. For example, most browsers require add-ons to be installed (Safari is an exception as it has close integration with Apple's Bonjour implementation of DNS-SD). Therefore the framework includes a web service that listens for DNS-SD announcements and forwards appropriate information to clients through a RESTful API.

7.2 Layers

For the IP Studio to be both useful and practical outside individual 'silos', its architecture has to balance robustness of real-time operations with the requirement to provide access from a range of different locations and on a range of platforms. The approach taken (Fig. 5) is to identify

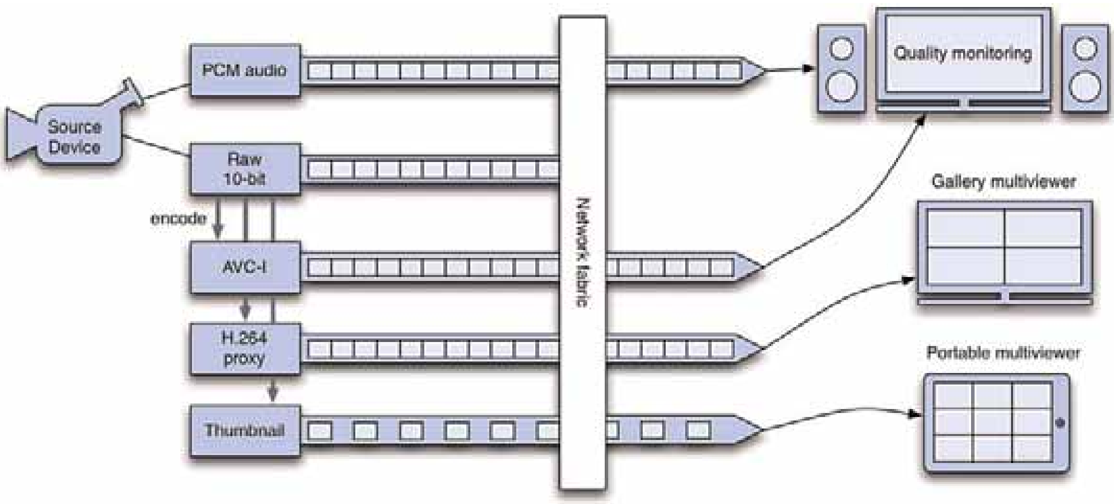


Figure 4 *Flexible viewing*

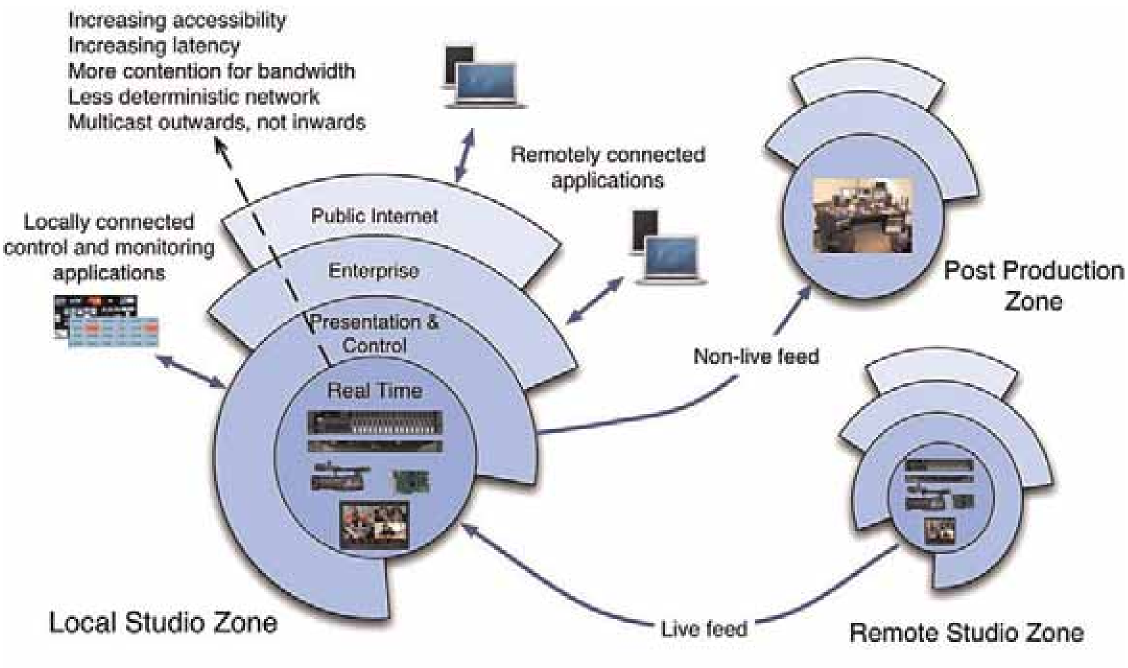


Figure 5 *Layers, zones and feeds*

different **layers** of activities carried out by a facility, with different requirements in terms of network characteristics and accessibility. So the most mission-critical real-time applications happen in the innermost layer, requiring a deterministic network and low-latency processing. Users working on a production over the Internet access an outer layer. Each layer stores grains that are appropriate to its activities, so the real-time layer provides the most recently recorded full quality content, while the outer layers serve web proxies.

Service gateways control what can be discovered outside a layer, and who/what can access it; this is essential to prevent connections from outer layers overwhelming the processing resources of inner layers. Similarly, forwarding of multicast messages is permitted from inner to outer layers, but not vice versa.

Expected network characteristics of different layers demand different approaches to transport of grains. Outer layers are likely to be less resilient and require use of TCP-based protocols to guarantee reliability.

The structure of a grain allows any combination of payload and transport protocol. This not only provides flexibility in facility configuration, but will also allow future formats and transports to be adopted with minimal disruption.

7.3 Zones and feeds

Many productions occur across multiple facilities, or **zones** in the project's terminology (Fig. 5). For example a local zone in a television centre might take contributions from a remote zone at an outside broadcast location. In the IP Studio, this is achieved by publishing **feeds** of sources from the remote zone, enabling them to be accessed in the local zone. Such a live feed is likely to require a low-latency stream at the real time layer, while in other cases, a feed might use a non-live transport from an outer layer, as is the case for the feed to post production in Fig. 5.

In some cases all the sources might be published, such as where the remote zone is being monitored and controlled from the local zone. In other cases, only some might be published, for example just the output of the vision mixer, and a filtered version of events.

7.4 Software defined networking

Successful scaling of the techniques discussed here will also depend on the network fabric that will take the place of SDI, AES and timecode distribution infrastructure in traditional facilities. Configuration of multicast group addresses and forwarding rules has proved quite time-consuming at BBC R&D's two demonstration zones at London and Salford, and a more automated approach will be needed for large installations, especially those that make use of dynamically provisioned networks. The project is investigating the use of software-defined networking (SDN) technologies, such as those being developed by the OpenFlow project [17].

8 Conclusions

BBC R&D's IP Studio project aims to provide programme makers and their audiences with a platform to create, consume and interact with new richer forms of content. This paper has outlined a framework to support flexible IP networking at the heart of the live production infrastructure. A unified approach to the identification and synchronisation of audio and video content and data and control events, and the adoption of techniques and technologies proven to scale for the Internet provide a level of flexibility that is not possible with traditional infrastructures.

A potentially disruptive finding is the proposal to distribute and consume content and events as elemental grains. While this enables great flexibility for a variety of application use cases, it is something of a step-change from the usual practice of layering traditional bitstreams onto ever-faster transports such as SDI over fibre.

A software framework has been implemented to test the feasibility and usefulness of this approach in practice, through trials with productions and infrastructure providers, and to determine where future standards are needed. A joint task force on networked media has recently been established by VSF, EBU and SMPTE [18].

9 References

[1] BBC R&D website, www.bbc.co.uk/rd/about/our-purpose

[2] BBC R&D IP Studio project, www.bbc.co.uk/rd/projects/ip-studio

[3] IEEE 1588, Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems

[4] RICHARDSON L., RUBY S.: 'RESTful Web Services' (O'Reilly, 2007)

[5] SMPTE ST 2022-6:2012, Transport of High Bit Rate Media Signals over IP Networks

[6] IETF RFC 3550, RTP: A Transport Protocol for Real-Time Applications

[7] IETF RFC 4607, Source-Specific Multicast for IP

[8] ISO/IEC DIS 23009, Dynamic Adaptive Streaming over HTTP (DASH)

[9] Pantos, R., 2013, IETF Internet-Draft, HTTP Live Streaming

[10] IETF RFC 6455, TheWebSocket Protocol

[11] øMQ: The Intelligent Transport Layer, www.zeromq.org

[12] SADALAGE P., FOWLER M.: 'NoSQL distilled: a brief guide to the emerging world of polyglot persistence' (Addison-Wesley, 2012)

[13] oRTP library, www.linphone.org/eng/documentation/dev/ortp.html

[14] SMPTE RP 2027:2012, AVC Intra-Frame Coding Specification for SSM Card Applications

[15] BBC R&D Stagebox project, www.bbc.co.uk/rd/projects/stagebox

[16] CHESHIRE S., KROCHMAL M.: 'IETF Internet-Draft: DNS-Based Service Discovery', 2011

[17] OpenFlow project, www.openflow.org

[18] VSF/EBU/SMTE Joint Task Force on Networked Media

Selected content from IBC2013

ISSN 2041-5923

Application of HEVC and DVB-T2 to terrestrial 4K UHDTV broadcast over-the-air trials

Sangjin Hahm   Zungkon Yim   Byungsun Kim   Sungho Jeon   Injoon Cho

Technical Research Institute, KBS (Korean Broadcasting System), South Korea

Abstract: The sense of presence in ultra high definition TV (UHDTV) mainly arises from a wide field of view combined with ultra high definition resolution and an extremely fine picture quality. Therefore, UHDTV requires a resolution of 4K (3840 × 2160) or 8K (7680 × 4320) pixels. Furthermore, to provide a more vivid picture, UHDTV goes as far as a frame rate of 120 fps, a wider colour gamut and a colour-depth of 12 bits. For this reason, UHDTV necessarily results in an extremely huge amount of data. Even many experts have questioned whether the 6MHz or 8MHz bandwidth of a terrestrial TV channel offers realistic possibilities for UHDTV transmission.

In Korea, the Korean Broadcasting System (KBS) carried out the world's first terrestrial 4K 30p UHDTV experimental broadcast from 9th October to 31st December 2012. The most prominent feature of this broadcasting was the use of a terrestrial TV channel. To fit the extremely high video data-rate into its limited bandwidth, we exploited high-efficiency video coding (HEVC) compression and DVB-T2 transmission and reception systems.

In this paper, we shall give an overview of KBS' 4K-UHDTV experimental broadcast and explain 4K-UHDTV content compression with an HEVC codec and the DVB-T2 system used in the trial. In addition, we shall show the result of our optimisation tests on HEVC coding of 4K 60p UHDTV.

1 Introduction

In Korea, on 31st December 2012, the termination of analogue terrestrial broadcasting was succeeded by a complete changeover to digital terrestrial high definition TV (HDTV). HDTV is no longer a new technology to the general public because its availability for over 10 years, has long since removed any novelty. Moreover, TV viewers are starting to request a post-HDTV broadcasting service and many market participants are already rapidly responding to this.

Now ultra high definition TV (UHDTV), together with 3DTV, is one of the highly anticipated successors in the post-HDTV era. UHDTV refers to a television system having four times or 16 times as many pixels, 3840 × 2160 or 7680 × 4320, as that of HDTV [1].

Unlike 3DTV which gives realism by binocular disparity, UHDTV covers a large part of the viewer's field of view and the resultant wide viewing angle encourages a sense of presence. For a faithful rendition of reality, UHDTV technology goes beyond a simple increase in screen size, however; everything is better in UHDTV. A frame rate of 120Hz and progressive scan contribute to the depiction of clear fast moving objects, 12 bits of colour-depth vivifies the subjects and 22.2-channel audio puts you right in the middle of the venue's sound field [1]. The sum of all these is a vast quantity of video/audio data and new technologies to deal with those must be developed and standardised.

In Korea, Korean Broadcasting System (KBS) conducted the world's first terrestrial 4K-UHDTV experimental broadcast from 9th October to 31st December 2012 employing DVB-T2 and high-efficiency video coding (HEVC) technology. The outline of the experimental broadcast is depicted in Fig. 1. In addition, KBS plans to

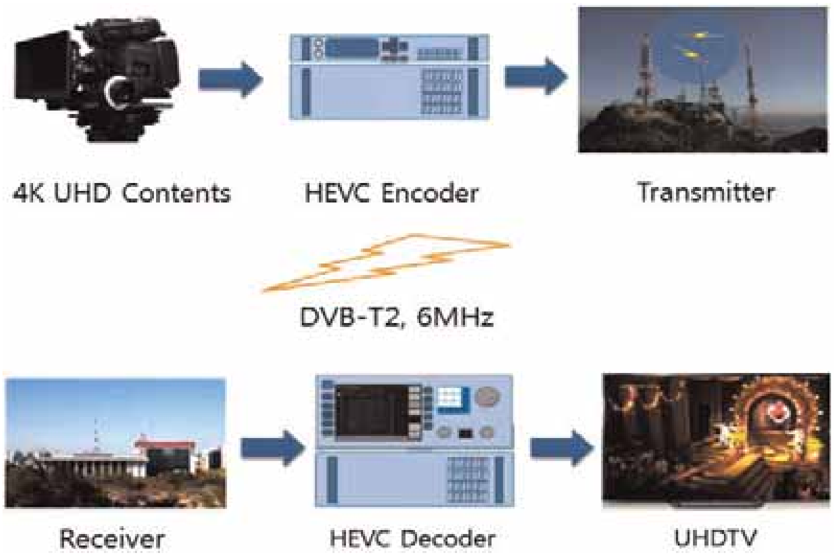


Figure 1 *KBS' terrestrial 4K-UHD experimental broadcasting*

execute experimental 4K 60p UHDTV broadcasting in the second half of this year.

In the following sections, details of KBS' test broadcast will be explained topic by topic. The first concerns the UHDTV video format, the second is about HEVC encoding of UHDTV content and the generation of its TS stream, the fourth describes the DVB-T2 transmission system. Finally, we explain the optimisation test results on 4K 60p HEVC encoding.

2 UHDTV video format

UHDTV video and audio signal formats are partially standardised but some sub-systems still remain undecided. Therefore, opportunities are still open for everyone. As already mentioned, the UHDTV video format includes 4K (3840 × 2160) and 8K (7680 × 4320) resolutions. The related standards are ITU-R BT.2020 (UHDTV parameters), BT.1201-1 (video resolution), BT.1361 (colorimetry), SMPTE 2036-1 (video signal), and SMPTE 2036-2 (audio signal). Table 1 summarises UHDTV video standards (parameters we chose for our experimental broadcasts are shown in red). Table 2 shows the UHDTV video data rates according to the optional system parameters [2].

As seen in the Table 1, the UHDTV standard includes various bits per pixel, chroma sampling and scanning methods. Therefore, the amount of data needed can vary from fourfold to 16-fold above HDTV, depending upon the exact parameter composition. Owing to the limitations imposed by the gap between the idealistic UHDTV specification and the current level of possible UHDTV implementation, we chose to adopt 4K, 30p, YUV 4:2:0, 8 bits (colour-depth), HEVC encoded video and stereo audio

Table 1 UHDTV video standard

	UHDTV		HDTV
	4K	8K	
Pixels	3840 × 2160	7680 × 4320	1920 × 1080
Frame rate	24p, 25p, 30p, 50p, 60p, 120p		60i
Pixel depth	10, 12 bits		8, 10 bits
Sampling	4:4:4, 4:2:2, 4:2:0		4:2:0
Aspect ratio	16:9		16:9
Audio channels	10.1–22.2 ch		Stereo, 5.1 ch
Field of view	55°	100°	30°
Viewing distance	1.5H	0.75H	3H
	H, the height of the screen		

Table 2 UHDTV video data rates

	Frame rate	Chroma sampling and colour depth	Data amount (Gbps)
HD	30	Y/CbCr 4:2:2 10 bits	1.16
4 K	30	Y/CbCr 4:2:2 10 bits	4.63
		Y/CbCr 4:2:2 12 bits	5.56
		RGB 4:4:4 10 bits	6.95
		RGB 4:4:4 12 bits	8.34
	60	Y/CbCr 4:2:2 10 bits	9.27
		Y/CbCr 4:2:2 12 bits	11.12
		RGB 4:4:4 10 bits	13.90
		RGB 4:4:4 12 bits	16.69
8 K	60	Y/CbCr 4:2:2 10 bits	37.08
		Y/CbCr 4:2:2 12 bits	44.49
		RGB 4:4:4 10 bits	55.62
		RGB 4:4:4 12 bits	66.74

to fit the UHDTV within the 6MHz bandwidth of current terrestrial digital broadcasting. The use of YUV 4:2:0 and a colour-depth of 8 bits was due to the limitation imposed by HEVC main profile.

3 4K 30p UHDTV experimental broadcasting

3.1 HEVC compression and creating a MPEG-2 TS

To deliver 4K content via a terrestrial channel within a 6MHz bandwidth, requires HEVC compression technology which can achieve 4 times the efficiency of MPEG-2 [3, 4]. HEVC is now under a standardisation process led by JCT-VC ('Joint Collaborative Team on Video Coding'). It was established by MPEG of ISO/IEC and VCEG of ITU-T in 2010 with the aim of developing a next-generation codec having more than twice the performance of H.264/AVC. The standardisation process of HEVC is scheduled to be completed in 2013 and it is considered to be the strongest contender for a UHDTV codec.

KBS is participating in a government-run R&D project for the development of a real-time HEVC encoder and decoder in collaboration with the Electronics and Telecommunications Research Institute (ETRI) and Kai Media Co. Ltd. and has applied the resultant decoder to the world's first terrestrial UHDTV experimental broadcast.

First, we encoded a YUV 4:2:0 30p file using the HM (HEVC Model) 6.0 standard software encoder with the parameters shown in Table 3. When we planned our 4K UHDTV broadcasting, the current version of HM was 6.0. The function of rate-control in HM 6.0 was not working well at that time, so we controlled the rate of the encoded bit stream by adjusting the quantisation parameter (QP) value.

We used MPEG2-TS as a transmission format for our broadcasts. In our multiplex, we combined the encoded video with an AC3 audio file.

3.2 UHDTV transmission and reception

The 4K UHDTV field trial transmission system was installed at the KBS Gwan-Ak station, one of the main commercial TV transmission sites covering the southern Seoul metropolitan area of South Korea. It is based on Digital Video Broadcasting – Second Generation Terrestrial (DVB-T2) standards with 6MHz bandwidth and 785MHz centre frequency and 100W transmission power. Using these transmission parameters, we conducted a field measurement campaign at 15 points having almost the same 5 km radius from the site. We tried to keep an equal angle interval for each measurement point.

DVB-T2 is the extension of the television standard DVB-T devised for the broadcast transmission of digital terrestrial television. DVB-T2 employs many advanced technologies such as low-density parity check (LDPC) code, a rotated constellation and 256-QAM symbol mapping with a large fast Fourier transform (FFT) size, so that an enhanced data capacity with high spectral efficiency can be achieved [5, 6].

Table 3 Encoder details

Video codec	HEVC (HM 6.0)
Input video	Size: 4 K (3840 × 2160)
	Sampling: YUV 4:2:0
	Colour depth: 8 bits
	Frame rate: 30 fps
Encoder setting	Main profile
	Maximum CU size: 64 × 64
	Intra-period: 32
	GOP size: 8
Rate control	Not available
	*Manually adjusted negotiating picture quality and bit rate
	Quantisation parameters used: 28, 30, 32, 34, ……
Bit rate	30 Mbps

3.3 Specification of the transmission system

The DVB-T2 transmission parameters for our experiment are listed in Table 4. We installed three panels (excepting 240°, so as not to interfere with an existing commercial DTV service using the same band). For stable reception, we set up the two dedicated antennas on the rooftop of the KBS laboratory building which is 10.59 km away from the Gwan-Ak transmission site and meets a line-of-sight (LoS) condition. The altitude of the reception antenna was 36 m.

3.4 Field trial parameters

In Table 5, common DVB-T2 system parameters are listed. The combination of parameters chosen was designed to cope with the data of a UHDTV service. In order to obtain a bit-rate of 30Mbps, the minimum required bit-rate, three FEC parameter sets were defined as Table 6. For experiments, the LDPC code rate was simply changed as defined.

The threshold of visibility (ToV) is measured as a function of two major measurement criteria: one is the carrier-to-noise ratio (CNR) value in the additive white Gaussian noise (AWGN) from laboratory tests and the other is the modulation error ratio (MER) value for the over-the-air signal received at the rooftop of KBS laboratory. The CNR and/or MER values were measured by applying 1dB step attenuation to the received on-air signal using a power attenuator.

3.5 Transmission test results

3.5.1 Rooftop fixed reception: If a parameter set 1–3 from Table 6 is applied to DVB-T2, a 4K-UHDTV channel can be supported even within a spectrum bandwidth of 6MHz. However, to obtain the minimum data rate of 35Mbps, the required CNR is around 22.4dB in AWGN. Also, a more strict MER of 24.7dB is required, even in the condition of LoS. Note that the official ToV CNR of ATSC-8-VSB is 15dB under AWGN. Specific values are shown in Table 7.

Table 4 DVB-T2 transmission parameter

Parameters	Description
Bandwidth	6 MHz
Centre frequency	785.000 MHz
Antenna power	100 W
Antenna panel	4 Dipole, 3 Panels 60/150/330
Altitude of antenna	690 m
Antenna gain	6.01 dBi
Polarisation	Horizontal

Table 5 Common DVC-T2 system parameter

Parameters	Description
FFT size	32 K mode
Guard interval	1/128
Pilot pattern	PP7
LDPC	64 800 blocks
Symbol mapper	256 QAM
Rotated constellation	On
Time interleaver/length	Single type / 3

Table 6 FEC parameter sets and effective data rate of 6 MHz bandwidth

FEC parameter set	FEC code rate	Effective data rate (Mbps)
1	5/6	36.568597
2	4/5	35.077556
3	3/4	32.873545

Table 7 Results of rooftop fixed reception

FEC parameter set	ToV@AWGN (CNR in dB)	ToV@KBS on-air (MER in dB)
1	20.8	24.3
2	22.4	24.7
3	23.0	25.1

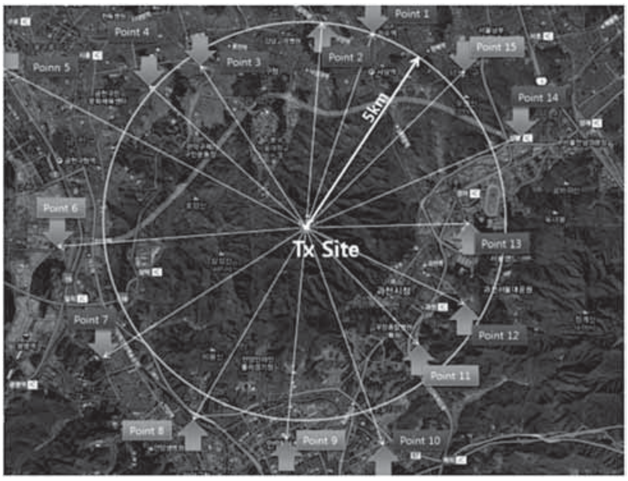


Figure 2 *Location of the 15 measuring points*
Each point is about 5 km distance from transmission site

3.5.2 Measurement systems and the results of the field trial: The system parameters used for the field trial are shown in Table 5 and the FEC parameter set 1 of Table 6 was used to maximise the capacity of transmission. Using these transmission parameters, we conducted the field test for 15 points with almost the same radial distance of 5 km from the transmission site. We tried to keep an equal angle interval for each measuring point as depicted in Fig. 2. The statistics of measured results are summarised in Table 8.

Measuring the received T2 signals required a test van carrying many devices to collect parameters such as: channel delay profiles, reception power levels, constellation diagrams, MER parameters, the geological land-shape profiles and the GPS data. In particular, the test van employs a 9.2 m-high mast system with remotely controlled rotor in order to receive signals under conditions similar to a domestic rooftop antenna.

From the analysis of each point, the places with LoS to the transmitter reveal a reception power level of around −50dBm and the other places with good reception (in spite of being non-LoS) are at a level of about −63dBm. Reception power levels are usually very low in areas where there is no LoS to the transmitter.

The average MER at the good reception points is 30.8dB which is about 4.6dB greater than the MER result of 25.1dB shown in Table 7 with parameter set 3. The MER values corresponding to bad reception points could not be recorded because the T2 measurement system could not achieve frame synchronisation.

In Fig. 3, the results of the field trial are depicted according to FEC parameter sets. For comparison, the CNR of ATSC under AWGN and a Brazil-A fading channel are also depicted. Compared with the AWGN case, the CNR and MER values are greater in the fading channel. Also, 4K-UHDTV transmission required a higher MER than that of ATSC.

4 Case study: application of HEVC TO 4K 60p UHDTV

For realistic UHDTV viewing, it is well known that it is necessary to increase the frame rate depending on the precise screen size. Therefore, we shall certainly need further experimental UHDTV broadcast at the rate of 60 frames per second with 4K resolution.

We plan to use HM 10 (Final Draft International Standard (FDIS)) as an encoder in our second experimental broadcast. The rate control function of HM version 10 is working well, unlike HM version 6.

Two experiments of HEVC encoding were performed for 60p UHDTV broadcast prior to the over-the-air trials. Fig. 4

Table 8 Received signal power

	Count	Average signal power (dBm)	Average MER (dB)
Good reception	10 points	−53.9	30.78
Bad reception	5 points	−73.8	Non-available

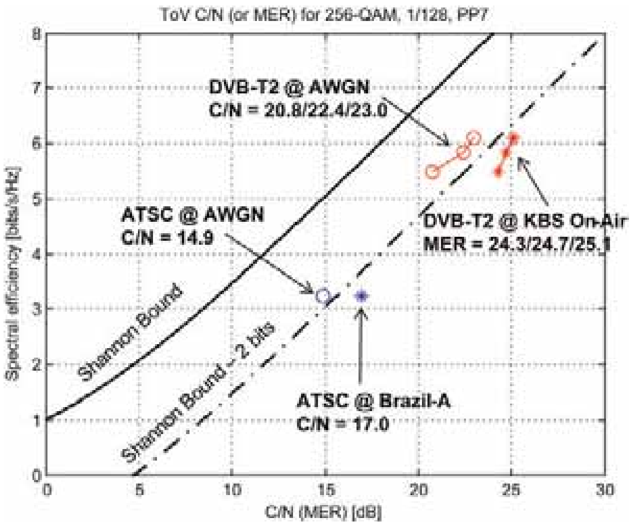


Figure 3 *Spectrum efficiency as a function of CNR or MER*

shows the sequences that were used in experiments. They were composed of the reference HD sequences used by JCT-VC and also of some of our own 4K video.

First, how many more bits are generated in encoding 60p UHDTV as compared with 30p UHDTV? The test of the increase in bit-rate was executed under the condition of the same QP value. Using the same QP value means a similar encoded video quality for each particular video sequence but with a different frame rate.

The result of the increase in bits with frame rate is shown in Table 9. It can be seen that bit-rate increases by approximately 17% to 40% between 30p and 60p.

Second, which configuration of intra-period and GOP (Group of Pictures) size is optimum in the HEVC encoding of 4K 60p UHDTV? Intra-period and GOP size

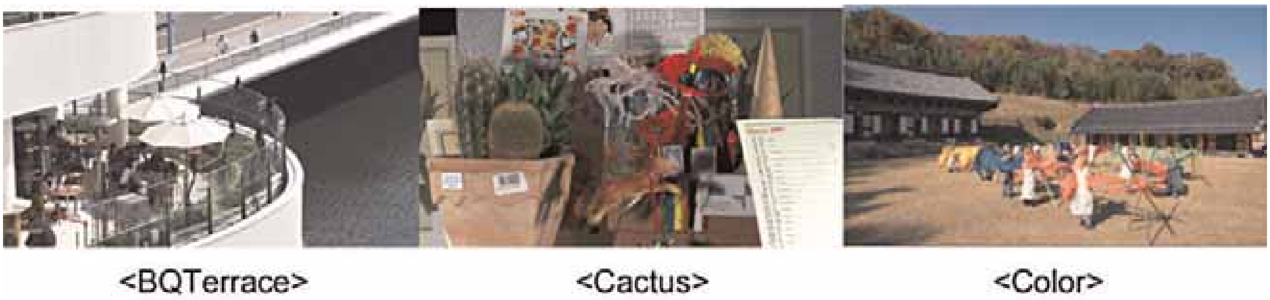


Figure 4 *UHDTV test sequences*

Table 9 Test result of changing in the frame rate

Test sequences	QP	30 P		60 P		BD	
		Bit-rate (kbps)	PSNR-Y	Bit-rate (kbps)	PSNR-Y	PSNR	Rate mode (%)
BQTerrace HD (1920 × 1080) JCT-VC	22	15 812.02	36.795	18 631.29	36.306	−0.2697	17.82
	27	3872.01	34.866	4205.54	34.744		
	32	1476.70	33.473	1567.59	33.439		
	37	719.79	31.760	756.81	31.762		
Cactus HD (1920 × 1080) JCT-VC	22	8933.35	38.051	10 383.02	37.854	−0.4222	18.22
	27	3394.45	36.315	3938.54	36.256		
	32	1614.56	34.210	1857.01	34.195		
	37	843.56	32.001	962.14	31.995		
Colour HD (3840 × 2160) KBS	22	106 840.81	39.609	148 722.41	38.996	−0.9227	32.25
	27	25 452.59	36.154	34 284.64	35.436		
	32	7408.75	34.852	9616.56	33.716		
	37	3022.75	33.475	3787.51	32.520		
Average						−0.5382	22.76

Table 10 Test result of changing GOP size and intra-period in 'Colour' sequence of 4 K 60 P

GOP	QP	Intra-period 32		Intra-period 64	
		Bit-rate (kbps)	PSNR-Y	Bit-rate (kbps)	PSNR-Y
8	22	148 722.41	38.996	133 652.57	38.167
	27	42 032.28	35.436	34 284.64	35.014
	32	9616.56	33.716	8054.66	33.482
	37	3787.51	32.520	3474.85	32.140
16	22	126 638.57	38.315	**113 974.98**	**37.917**
	27	29 569.60	34.921	**24 749.61**	**34.677**
	32	9589.23	33.434	**6668.89**	**33.274**
	37	2985.04	32.343	**2645.13**	**32.205**
32	22	100 942.08	37.099	91 524.80	37.248
	27	18 755.22	34.482	16 263.63	34.481
	32	5205.10	33.224	4414.78	33.201
	37	2411.89	32.168	1788.36	32.125

influence both the video quality and the amount of bits generated by HEVC encoder. If a smaller intra-period is used in an encoder, there is the advantage of smaller random-access interval in the encoded video and a smaller channel switching delay time when watching TV. However, there is a drawback in an increase of the amount of bits encoded by HEVC. In the case of GOP size, using a larger GOP size increases the memory usage and processing delay when encoding and decoding, but has the advantage of increasing the compression ratio. Therefore, experiments were carried out to set intra-period and GOP size and Table 10 shows the result of the test. According to our experimental results, the optimal intra-period arises as 64 and optimal GOP size emerges as 16, to encode 4K 60p UHDTV by HEVC.

5 Conclusion

UHDTV and 3DTV seem to be the most prominent successors of the post-HDTV era. The relatively low technological difficulty of UHDTV made the broadcast industry of Korea move promptly toward it. Now, 4K displays are rushing to the market in search of a new source of revenue and KBS, as the public broadcaster responsible for establishing future broadcasting for the good of national and public wealth, has a strong will to carry forward terrestrial UHDTV broadcasting experiments in a bid to invigorate UHDTV broadcasting.

To demonstrate this willingness, KBS worked on 4K-UHDTV (30p) experimental broadcasts last year and plans to carry out 4K-UHDTV (60p) over-the-air trial soon. After the completion of this experiment, KBS will ceaselessly devote its best efforts to fulfilling its aspiration of launching 4K-UHDTV services in 2018 and 8K-UHDTV experimental broadcasts for the beginning of the Pyungchang 2018 Winter Olympic Games.

6 References

[1] ITU-R BT. 2020:2012. Parameter values for UHDTV systems for production and international programme Exchange

[2] SMPTE 2036-3:2010. Ultra High Definition Television – Mapping into Single-link or Multi-link 10 Gb/s Serial Signal/ Data Interface

[3] JCTVC-H0022r1. JCT-VC AHG report: HM subjective quality investigation (AHG22), Feb, 2012

[4] JCTVC-H0360r1. Comparison of Compression Performance of HEVC Working Draft 5 with AVC High Profile, Feb, 2012

[5] EN 302 755 V1.3.1, Frame Structure Channel Coding and Modulation for Second Generation Digital Terrestrial Television Broadcasting System (DVB-T2), April 2012

Selected content from IBC2013

ISSN 2041-5923

Investigations into the characteristics of technologies for TV white space applications

T. Harrold   M. Waddell

BBC, UK

Abstract: The regulatory framework for access to the TV white space (TVWS) spectrum has been progressing rapidly in the UK and licence exempt access through a geolocation database should soon be available [1]. The characteristics of the white space devices (WSDs) that will exploit this spectrum are less well known, but a number of candidate technologies are emerging. The BBC has participated in two field trials in rural and sub-urban environments to evaluate the performance of three WSD candidate technologies. The range and throughput achieved for broadband applications in these UK trials is discussed. The selectivity and overload characteristics of TV receivers will play an important part in determining the power that can be safely radiated without disrupting TV reception. An extensive programme of laboratory measurements on TV receivers has been completed and reveals a considerable spread in performance. Some recent designs are proving particularly susceptible to TVWS interference and new performance targets are recommended to unlock the full potential of new TVWS services.

1 Introduction

To investigate the potential value of new TV white space (TVWS) devices a number of field trials have taken place in the UK. A large trial in Cambridge assessed the performance of a low-power wireless broadband base station using COFDM technology. A further set of trials on the Isle of Bute in Scotland considered two alternative technologies to implement broadband links in a more challenging rural environment. Results from these trials are presented.

The availability of TVWS spectrum is influenced greatly by the interference handling characteristics of the TV receivers that will share the spectrum. This will determine the permitted power and the range of applications that can be supported. An extensive programme of tests has been conducted to assess the interference characteristics of six candidate white space device (WSD) technologies. Interesting trends are emerging with certain TV receiver designs showing poor resilience to some TVWS signals. This area will require attention from receiver designers to prevent interference and increase the usefulness of the TVWS band.

2 The Cambridge white space trial

The trial in Cambridge brought together a diverse range of stakeholders in a consortium led by Microsoft [2]. Cambridge has strong Digital Terrestrial Television (DTT) coverage from a single dominant transmitter at Sandy Heath and so proved an ideal location for TVWS trials with 17 channels available for the trial, at up to 10W EIRP.

The BBC worked with partners from Arqiva and Adaptrum on the evaluation of a 125mW (21dBm) broadband access point using SISO COFDM technology [2]. The experimental modem was evaluated in the laboratory to determine the network throughput as a function of the received signal power and the characteristic is shown in Fig. 1. The modem shows a well-behaved characteristic, switching to higher-order modulation modes supporting greater throughput as the available signal to noise ratio increased.

The quality of TVWS spectrum is typically degraded by interference from distant DTT stations and the performance of the modem was evaluated operating on two TV channels to explore this effect. Channel 59 was subject

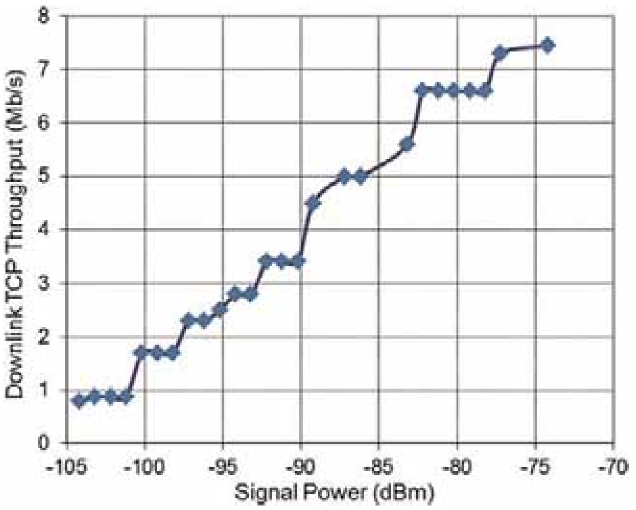


Figure 1 *Throughput versus power for the Cambridge modem*

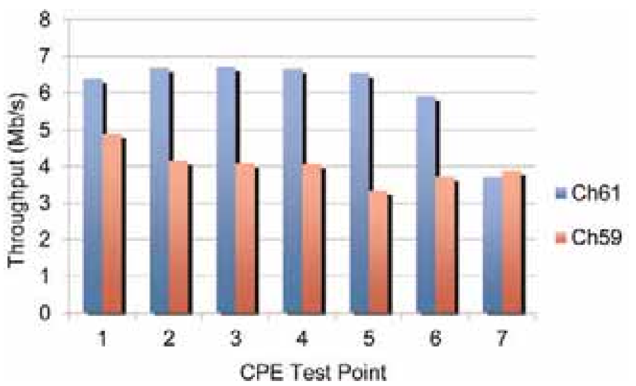


Figure 2 *Effect of DTT interference on throughput*

to co-channel interference from the Tacolneston DTT transmitter, while channel 61 was a relatively clean channel. Fig. 2 shows the difference in measured uplink throughput for the two channels. The cleaner channel (channel 61), which was found to have 6dB less co-channel interference from distant DTT services, shows a higher throughput.

The differences are also apparent in the coverage predictions for the base station operating in the two channels taking account of the DTT interference expected in the network. Fig. 3 shows the coverage predictions expressed in terms of the ratio between the downlink power and the received noise and interference (C/(I + N)). The reduced cell size for the TVWS base station operating in the noisier channel (59) is clearly apparent. Computer predictions and spectrum analyser measurements at the base station confirmed that channel 61 was some 6dB cleaner than channel 59.

3 The Bute white space trial

To assess the performance of TVWS links in a more challenging rural environment, BT installed equipment at a telephone exchange on the Isle of Bute in Scotland. Unlike Cambridge, there is no single dominant DTT transmitter for Bute and the island is at the edge of service receiving its DTT from up to 11 transmitters depending upon location. Only three TVWS channels could be made available by Ofcom and channel 57 (762MHz) and channel 52 (722MHz) were selected to evaluate two TVWS technologies. Analysis of DTT predictions revealed that channel 52 was a relatively clean channel and it was used to support a WiMAX network comprising a base station (BS) and four sets of customer premises equipment (CPEs). Channel 57, was subject to co-channel DTT interference

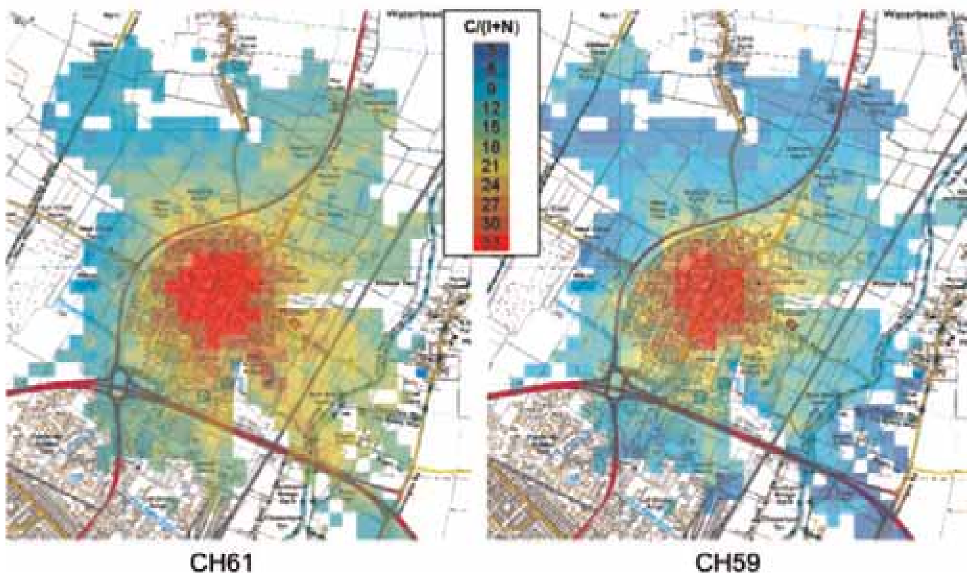


Figure 3 *Link coverage predictions for channels 61 and 59*

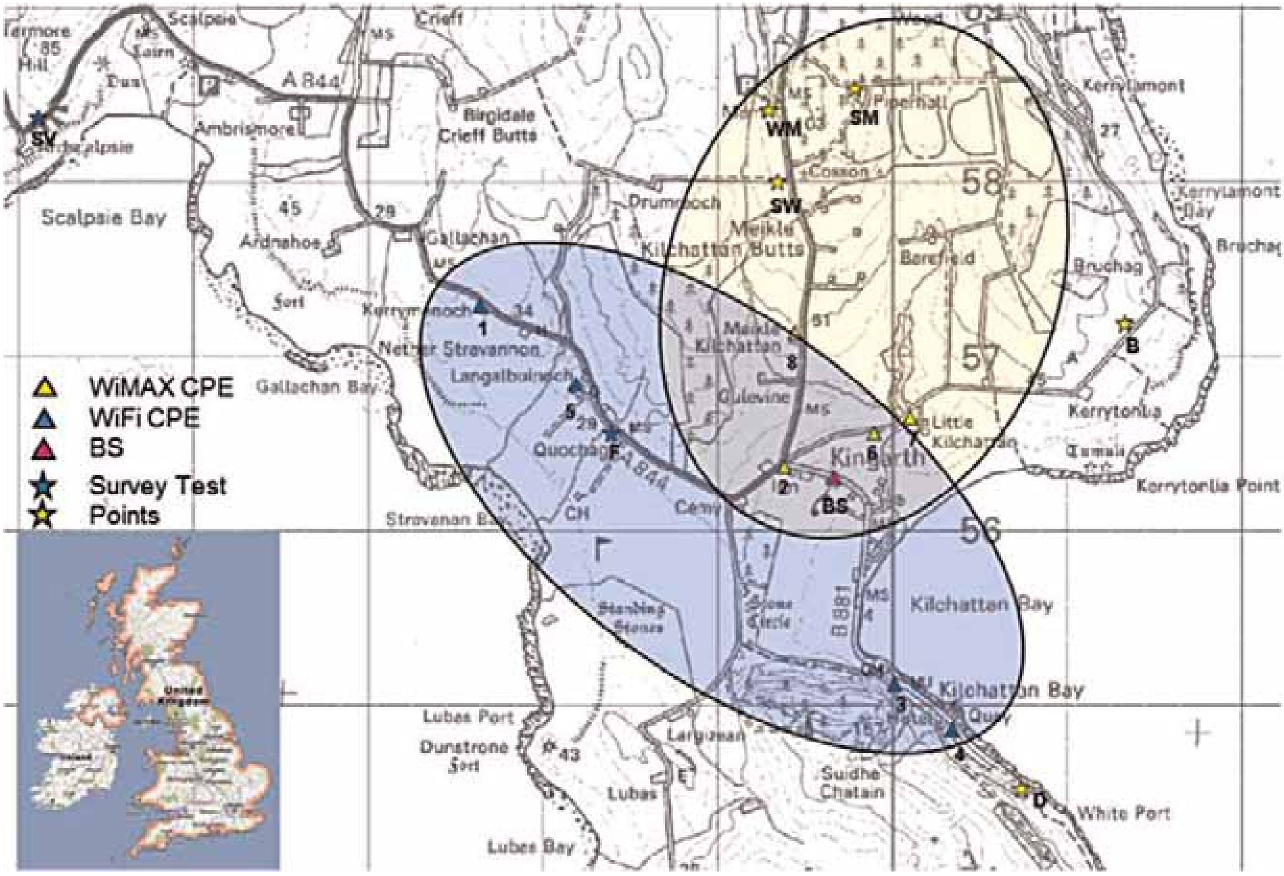


Figure 4 *Location of Bute CPE and BS deployments*

from Rosneath and South Knapdale, which serve part of the island, and was selected for a WiFi network with four additional CPEs. Directional antennas were used allowing higher radiated power than that used for the Cambridge trial. Both BS were operated at approximately 36dBm (4W) EIRP.

The location of the WiFi and WiMAX CPEs and the BS are shown in Fig. 4. The WiFi BS antenna was pointed to the northwest with a back lobe providing additional coverage to the southeast. The WiMAX BS antenna was pointed north-northeast.

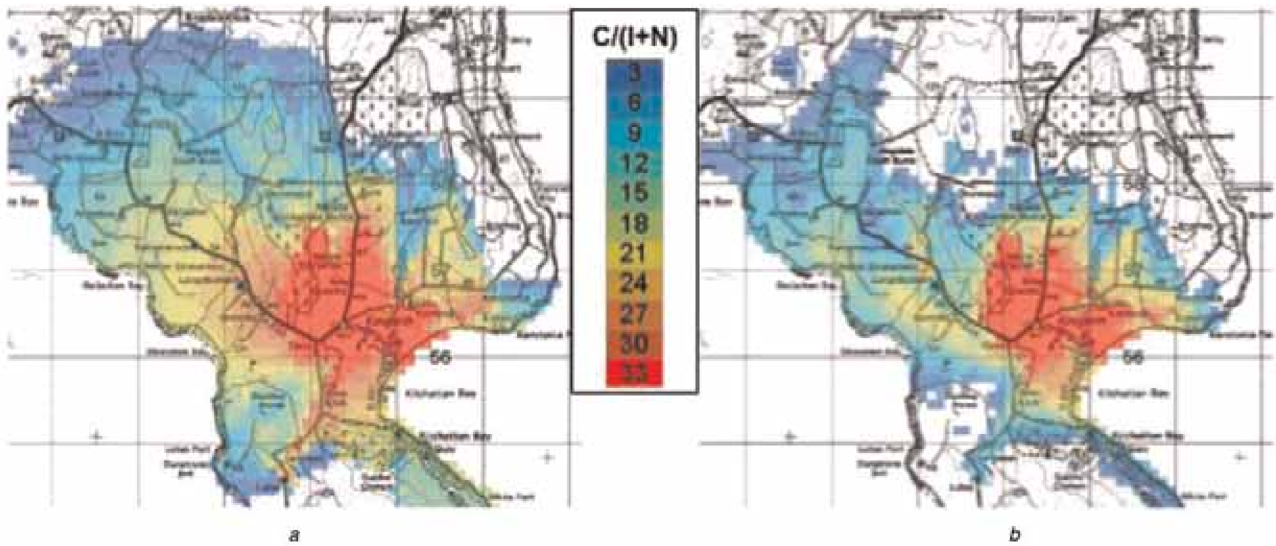


Figure 5 *Reduction in CH57 WiFi BS coverage (35 dBm EIRP) when 50% time DTT interference is considered*
a CH57 coverage – excluding DTT interference
b CH57 coverage – with DTT interference

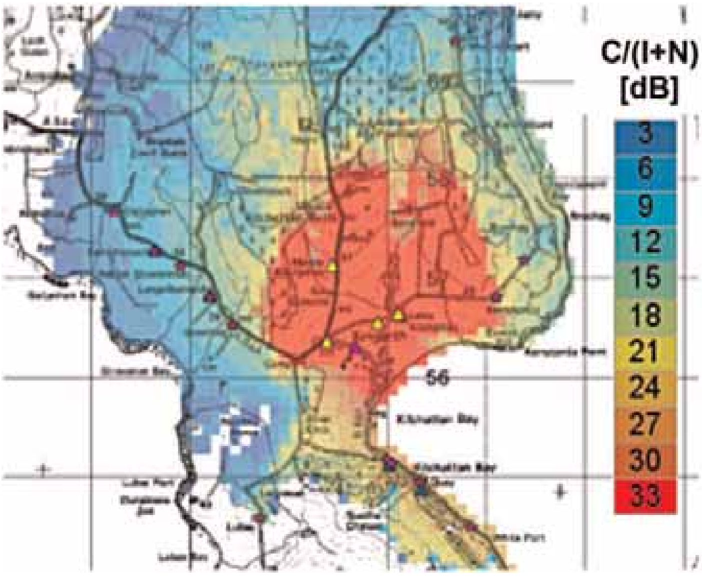


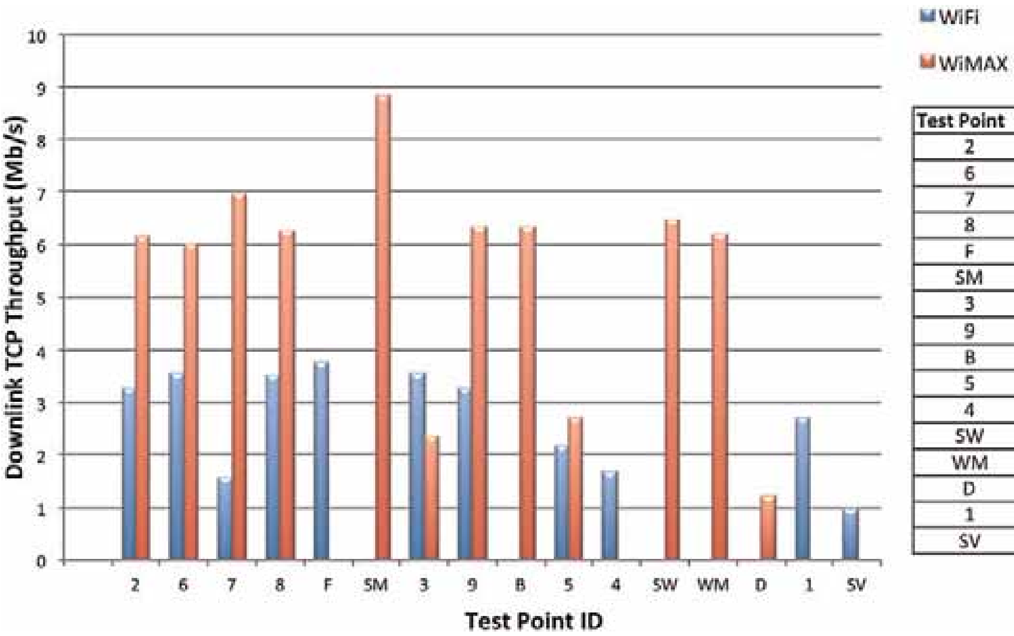
Figure 6 *CH 52 WiMAX coverage prediction accounting for 50% time DTT interference*

The impact of the co-channel DTT interference from Rosneath and South Knapdale is shown in the coverage predictions (Fig. 5). The left hand prediction shows the coverage accounting only for noise while the right-hand figure shows the reduction in area when the interference from Rosneath and South Knapdale DTT stations is considered in the simulation.

South Knapdale illuminates the south west of the island and Rosneath is to the north east. The CPE sites were chosen to minimise the effect of this interference.

The WiMAX channels were free from DTT interference and the predicted coverage for the 36dBm base station is shown in Fig. 6.

Analysis of the TCP downlink throughput at a number of test points revealed that the WiMAX network consistently outperformed the WiFi network. This is anticipated given the use of MIMO techniques and a more efficient co-ordinated MAC layer. The chosen test points are shown in Fig. 4 and the measured throughput rates are shown in Fig. 7. The test points are ordered according to range and



Test Point	Distance (m)
2	296
6	337
7	555
8	861
F	1228
SM	1228
3	1234
9	1516
B	1516
5	1573
4	1596
SW	1721
WM	2137
D	2139
1	2259
SV	4939

Figure 7 *Throughput comparisons for Bute test sites*

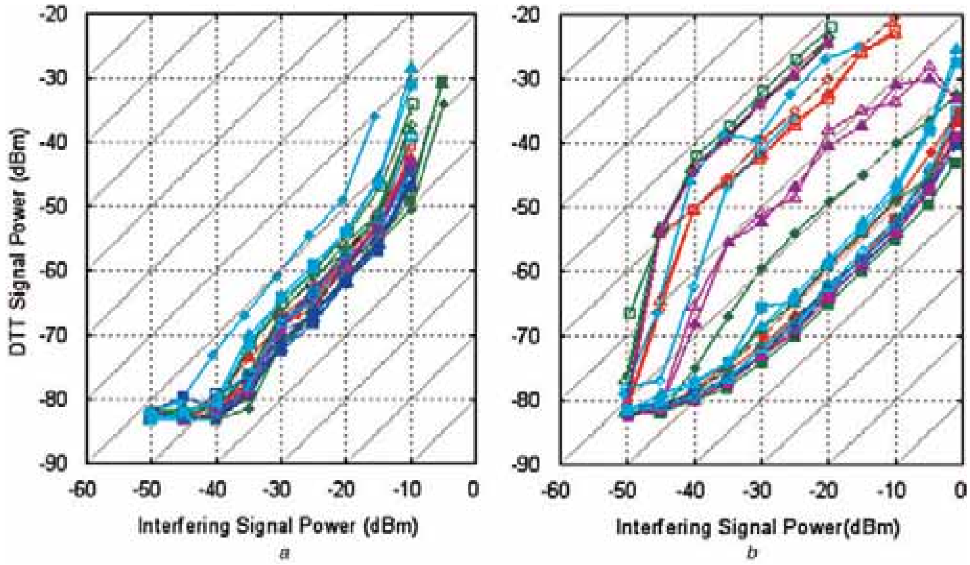


Figure 8 *Performance comparison for two receivers subjected to adjacent channel WSD interference from a range of WSD types*
a Good receiver, tolerant to TVWS
b Receiver intolerant to WSD interference

it can be seen that the WiFi network achieves a greater range albeit at a lower throughput.

4 DTT receiver tests

TV reception can be affected by the presence of white space signals in a number of ways. The receiver selectivity, AGC behaviour and overload characteristics will determine the minimum permitted ratio between the wanted DTT signal and the interfering WSD signal for stable reception, known as the protection ratio. This, together with the coupling loss between the WSD and the DTT antenna, will determine the maximum permitted WSD EIRP. The most likely interference mechanism is where signals from WSD CPE couple into neighbouring DTT installations, causing DTT picture break-up if the EIRP is not appropriately limited. Tests have been carried out on typical DTT receivers, in order to investigate the protection ratio required to prevent WSD interference.

4.1 Protection ratio measurements

Fourteen DTT receivers were evaluated for resilience against interference from six candidate White Space technologies. The DTT receivers included set-top boxes (STBs), integrated digital televisions (IDTVs) and personal video recorders (PVRs). Measurements revealed that half of the receivers used can-type (superheterodyne) tuners and the remainder used silicon (direct conversion) tuners. The interfering signals were generated from a vector signal generator replicating interference from proposed WSD technologies; both manufacturer-proprietary waveforms and established radio standards (WiMAX, WiFi and LTE) were used. Waveforms representing various traffic levels from the White Space technologies were used for both uplink and downlink.

A wide variation in performance was observed among the 14 receivers. Fig. 8 shows the performance of two DTT receivers in the presence of various types of closest adjacent-channel white space interference. The amount of DTT signal needed to decode a picture in the presence of a given level of WSD interfering power is plotted.

The performance of the receiver in Fig. 8*a* is relatively consistent for various types of white space interference. However, the receiver in Fig. 8*b* shows wide variation in performance in the presence of different types of white space interference. The results suggest that in most

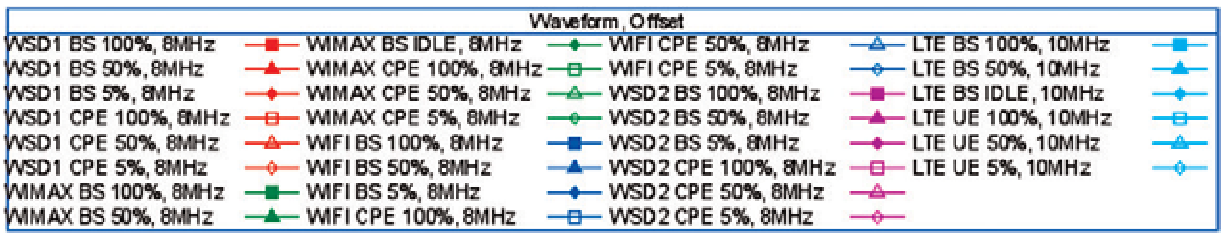


Figure 9 *Waveforms used for receiver tests*

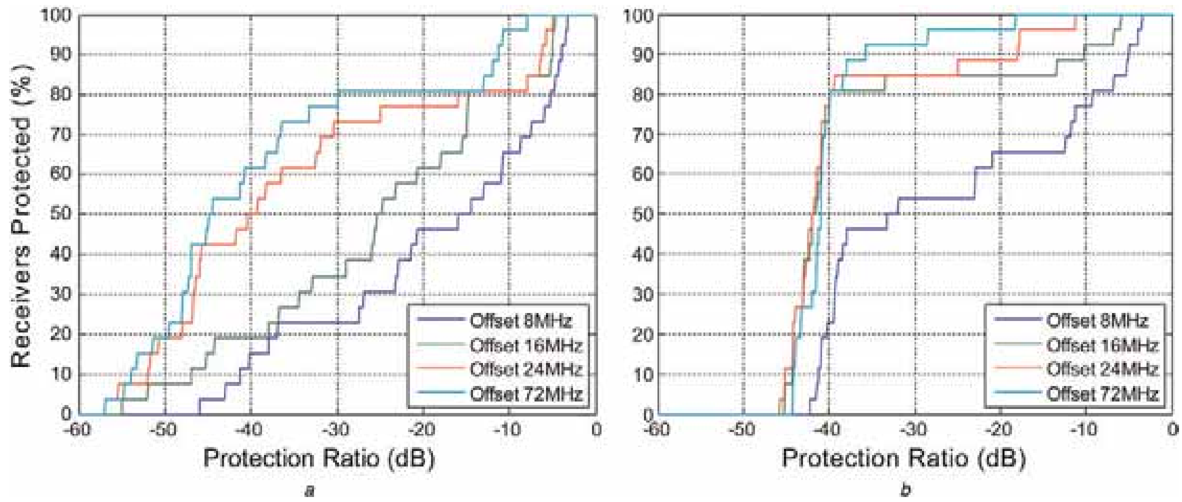


Figure 10 *Receivers protected for a WSD Signal Power of*
a − 20 dBm
b − 40 dBm

cases, i.e. in situations where the received DTT signal power is at its weakest (−80dBm), reliable DTT reception will be possible provided that the received adjacent-channel interference power is no greater than −45dBm.

The waveforms that caused the most interference to the intolerant receiver are characterised by having a 'pulsed' nature, with frequent 'on to off' transitions and a high peak to mean ratio. Such waveforms seem to adversely affect the operation of the automatic gain control (AGC) and channel state process in certain DTT receivers. Further investigations showed that the susceptibility to low-duty pulsed interference occurs even at large frequencies offsets (>72MHz), indicating that some designs have little front−end selectivity. The worst performers were all recent designs using a silicon tuner, although it should be noted that most silicon tuners have good performance.

4.2 Investigations with a pulsed waveform

Further investigations were carried out on a larger set of 26 receivers using one of the pulsed waveforms that had been found to provoke problems in the initial tests. The intention was to discover how widespread the vulnerability might be.

Fig. 10 shows the percentage of receivers from the set of 26 tested, that achieved a particular protection ratio. Results for WSD signal powers of −20 and −40dBm are shown, using a waveform generated by a WiMAX CPE operating at a low traffic level. Results such as these can be used by regulators to determine the protection ratios necessary for various offsets based on a requirement to protect a particular percentage of the receiver population. Given the significant spread in performance it is felt that test waveforms and performance targets should be developed to assist tuner manufacturers in the development of new designs.

The WiMAX waveform used in these tests is known to provoke AGC problems in certain vulnerable receivers and reveals performance deficiencies which are not apparent when considering interference from broadcast-like signals. It may be possible to develop future WSD technologies to avoid this problem but measurements using simple switched waveforms indicate this may be difficult. The vulnerable receivers exhibit similar weakness in the presence of the 'off to on' transition when an otherwise benign signal is switched on. Some of the proposed applications for WSDs include machine-to-machine communications (e.g. smart metering) and low demand web access. Such applications, where the signal is likely to appear intermittently for short periods, are likely to provoke sporadic picture failure as the AGC adapts to the interference.

5 Conclusions

Field measurements in Cambridge and the Isle of Bute have illustrated the potential value of new WSD technologies for broadband connectivity. The technology could prove particularly useful where conventional cabled technologies are uneconomic to deploy. For a 4W TVWS base station, fixed connectivity at a range of up to 5km has been demonstrated. Interference from existing DTT networks is likely to degrade performance in some channels and careful radio planning will be required to integrate such links with the existing DTT services.

The range of TVWS applications that can be supported by the new technologies will be dependent on the performance of the incumbent DTT receivers and tests have revealed issues with certain new silicon designs. Low duty cycle uplink signals can generate 20dB or more interference when compared to steady-state broadcast signals. This would appear to suggest that higher power applications should be confined to down-link only signals. Improved receiver design and less interfering WSD technologies will be required to unlock the full potential of the TV white space.

6 Acknowledgments

The authors would like to thank S. Thilakawardana and other colleagues at the BBC for their assistance with the field trials. We would also like to thank the project partners in the Cambridge and Bute Consortia and the Technology Strategy Board for funding the Bute project.

7 References

[1] OFCOM: 'TV white spaces: a consultation on white space device requirements' 2012

[2] CAMBRIDGE WHITE SPACES CONSORTIUM: 'Cambridge TV white spaces trial: a summary of the technical findings' 2012

[3] WADDELL M., THILAKAWARDANA S., HARROLD T., KESBY P., CHERRY S.: 'Performance of an experimental TV white space base station for mobile and fixed broadband applications' (BBC R&D White Paper WHP 228, 2013)

Selected content from IBC2013

ISSN 2041-5923

High dynamic range video cameras based on single shot non-regular sampling

Joachim Keinert   Michael Schöberl   Matthias Ziegler
Frederik Zilly   Siegfried Foessel
Fraunhofer Institute for Integrated Circuits IIS, Germany

Abstract: Compared with traditional video acquisition, high dynamic range video increases the visual experience, enhances opportunities in post-production and simplifies studio lighting. However, existing techniques for high dynamic range video capture are too expensive for low and medium budget cameras. Furthermore, they are not suitable for handling the most challenging lighting conditions. Against this background, this paper presents a novel technique for capturing high dynamic range video. It is based on a non-regular mask of attenuation filters placed in front of a standard sensor. Compared to a traditional sensor with the same resolution, the captured image shows nearly the same amount of detail, but has a much higher dynamic range. The achievable dynamic range increment depends on the underlying sensor capabilities. For highest quality requirements, a 12 f-stop sensor can be extended by 2 to 5 f-stops and a 14 f-stop sensor, by 4, to 7 f-stops. The obviation of bracketing avoids motion artefacts and reduces the data volume read from the sensor.

1 Introduction

The visual impression of video depends on many parameters such as recognisable details, frame-rate, representable colours and its dynamic range. The latter is defined as the ratio between the value of the brightest and the darkest recognisable pixel. While sensors with 4 k, or even 8 k, resolution allow the capture of very fine-grained details and high frame-rate cameras permit the reproduction of very smooth motion, it remains difficult to handle the complete dynamic range of a scene. McCann and Rizzi [1], for instance, report natural scenes with 14.3 f-stops and test charts with 17.9 f-stops. High-end cameras, on the other hand, provide a dynamic range of approximately 14 f-stops. Unfortunately, the underlying techniques are too complex for small-sized cameras and too cost-intensive for low and medium budget cameras, so these remain incapable of capturing the complete dynamic range of a scene. Moreover, even the best cameras do not provide enough dynamic range to handle challenging light conditions.

Nevertheless, capturing the full dynamic range of a scene is a highly desirable goal, since studies confirm that a high dynamic range is very beneficial for the visual experience [2] of the image. Furthermore, it increases opportunities in post-production and allows a reduction in the effort spent on studio lighting.

Against this background, the following sections describe a novel technology for the generation of high dynamic range (HDR) video. It is based on computational imaging principles that are able to increase the dynamic range of a sensor technology without impairing the visual quality of the image. In more detail, the paper reviews state-of-the-art HDR capture technologies, analyses their limitations and shows how these can be resolved by our new method. Results from a first camera prototype, and an analysis of the achievable increase in dynamic range, complete this paper.

2 State of the art

To fulfil the need for an increased dynamic range, sensor manufacturers have been improving their technology. Development toward an increased dynamic range, however, requires large pixels and comes at a high sensor cost. Alternatively, the performance of existing sensor designs can be extended with computational imaging methods.

For extending the dynamic range of an existing camera, many computational approaches have been proposed. In a traditional camera, a single exposure delivers the final image right away. In contrast, in computational photography, some form of intermediate representation is captured and the final image requires additional post-processing of the data.

The most basic method for capturing high dynamic range images is based on multiple sequential exposures of the same scene [3]. Subsequent processing combines these images into a single radiance map that directly represents the brightness of the scene [4]. This method is popular in still photography and can already be found in many consumer still image cameras (and even smartphones). Results come at the cost of higher rate processing within the camera. For still imaging, this is not a problem but a video camera needs to work at a high rate to cope with the additional exposures captured.

Moreover, the multiple exposure approach is hard to apply when the scene contains motion, as sequential exposure will lead to distortion. The sequential images will obviously not match, so an additional motion-compensation step will be necessary. But this merging might well fail, as is demonstrated in Fig. 1. The image shows a moving scene captured with a professional cinema camera. Due to the multiple exposure 'HDR mode', significant distortions are visible at the edges of moving objects.

In addition to non-matching object positions, different amounts of blur are also a challenge: the images vary significantly in the amount of blur caused by large differences in exposure time. Fig. 2 illustrates this: The left image has the correct amount of blur, as set by the exposure time. For recovering the overexposed details, a shorter exposure (centre) is made. It has significantly lower blur, which makes rolling shutter artefacts (like the distorted shape of the fan blades) much more visible. The fused image on the right mainly uses the short exposure component for recovering the fan blades. The HDR component introduces stuttering image content and rolling shutter artefacts into the final image, resulting in an unnatural look.

As this example illustrates, in critical scenes severely visible distortions can be caused which are not acceptable in professional applications. So, HDR video capture based on

Figure 1 *Multiple-exposure HDR: Similar scenes were captured with single-shot LDR mode (left) and multiple-exposure HDR mode (right)*

The latter shows severe distortion

Figure 2 *Multiple-exposure HDR: Long exposure (left) and short exposure (centre) are fused (right)*

Significant distortions are visible, while an all-blurry fan is expected

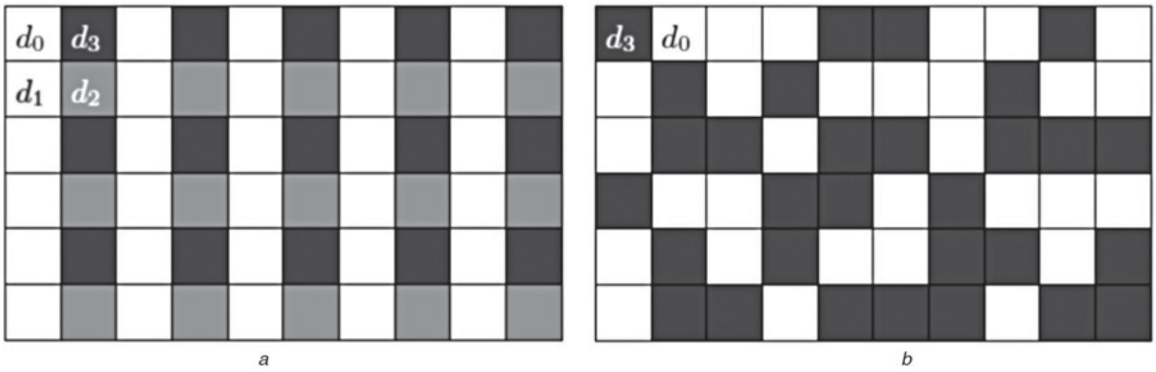


Figure 3 *Single sensor, single exposure HDR imaging with regular attenuation pattern [7] and proposed non-regular attenuation pattern*

a Regular attenuation pattern
b Non-regular attenuation pattern

multiple exposures does not represent an optimum solution. It would be especially bad for sports content, for example.

To circumvent these problems, both the bright and dark parts need to be captured simultaneously with the same exposure time. This can be achieved by using multiple synchronised cameras and optical filters. The most obvious approach is to use beam splitters or semi-transparent mirrors [5]. This poses alignment challenges similar to those of stereoscopic camera rigs and results in expensive, heavy and large setups that require accurate alignment. Another approach, based on a side-by-side array of cameras, avoids a complicated setup and compensates for the disparity algorithmically [6]. However, this approach can fail for certain scenes where obscuring objects cause both cameras to see different scene content.

Yet another approach to high dynamic range video imaging is available as a single sensor with spatially varying exposures [7]. As shown in Fig. 3*a*, a regular pattern of attenuation filters is placed over the pixels. This enables the sensor to capture both the bright and dark regions of the scene in a single exposure. As all image data is captured at once in a single sensor, moving objects are represented accurately and alignment is perfect. Unfortunately, this approach loses resolution due to the regular pattern.

3 Non-regular sampling

In contrast to a regular pattern, our mask of attenuation filters is based on a non-regular pattern. An exemplary section of the sensor is shown in Fig. 3*b*. The attenuated pixels capture image information in bright parts of the scene whereas the pixels without attenuation capture the darker parts. This leads to a non-regular sub-sampling for bright and dark image regions. With our specially-adapted image reconstruction algorithm a much higher resolution can be recovered from this data. In the following, the reasons for the non-regular arrangement and the resulting image reconstruction are discussed.

3.1 Non-regular sub-sampling principle

The non-regular attenuation pattern makes our technology fundamentally different from existing regular attenuation patterns [7]. In classical sampling theory, sub-sampling requires optical low pass filtering prior to sampling to avoid aliasing. For example, by sub-sampling by a factor of two, at least half of the resolution gets lost.

With the proposed non-regular sampling approach, we are able to utilise an important property of natural images: through transformation to a different representation (like Fourier or discreet cosine transform (DCT), only few coefficients are sufficient for representing the signal well. This so-called *sparsity* property is widely used in transform-based image compression. In the context of sub-sampling it provides a mathematical theory for being able to reconstruct the image signal without measuring all of the pixels.

As discussed in [8], the sampling of the original signal is still subject to aliasing. However, a suitable image reconstruction is able to identify the dominant signal components and reconstruct the image. By these means a convincing image quality can be obtained.

3.2 Resolution analysis

The proposed method is able to recover some of the resolution lost through sub-sampling, however, not all image details can be recovered perfectly, as some information has obviously been discarded. The quantity of the resulting loss depends on the image content. Consequently, resolution analysis is only possible for a specific image.

In order to evaluate the capability of our method to recover high-frequency details, simulations with a zone plate have been performed [9]. 25% of the pixels are supposed to be known, allowing thus for 4 different attenuation filters. As shown in Fig. 4, regular sub-sampling leads to severe

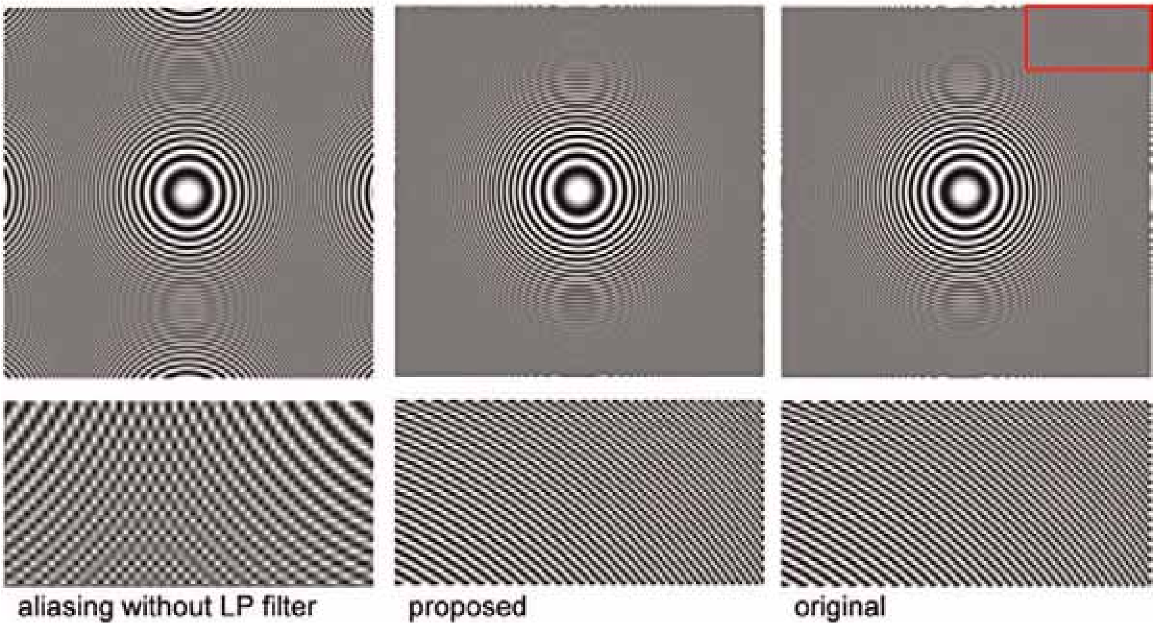


Figure 4 *Resolution recovery: Aliasing is visible for regular 25% sub-sampling and reconstruction (left), the proposed non-regular sub-sampling and reconstruction (centre) is indistinguishable from the original image (right)*

aliasing that cannot be removed with interpolation. For the proposed sub-sampling and reconstruction, the test chart is recovered in full detail without any visual difference from the original image. This shows that a perfect image reconstruction is possible. However, this is not universal result, and for a noise-like image structure, it is not always possible to obtain a perfect reconstruction. Luckily, we found this problem to be well-behaved; for typical natural images the resulting images are of high visual quality.

3.3 Theoretical increase in dynamic range for high-quality acquisition

Non-regular sampling permits an increase in the dynamic range of an image sensor while keeping high visual quality. This section aims to quantify the achievable increase in dynamic range.

The dynamic range of a camera can be derived from a plot that shows its signal-to-noise ratio (SNR) as a function of the input illumination. According to the EMVA Standard 1288 [10], the SNR can be computed as follows:

$$\begin{aligned} \mathrm{SNR}(\mu_\mathrm{p}) &= \frac{\eta \cdot \mu_\mathrm{p}}{\sqrt{\sigma^2 + \eta \cdot \mu_\mathrm{p}}}, \\ \mathrm{SNR}_\mathrm{dB}(\mu_\mathrm{p}) &= 20 \log_{10}(\mathrm{SNR}(\mu_\mathrm{p})) \end{aligned} \tag{1}$$

μ_p is the number of photons hitting the area of a single pixel during the exposure time. It hence represents the illumination of the pixel. η is the total quantum efficiency of the sensor indicating what fraction of the incoming photons can be transformed into electrical charges. σ^2 is a noise term and accounts for the sensor read-out and amplifier noise as well as quantisation noise.

Fig. 5 is a sketch plot of (1). The left curve represents the unmodified sensor without any mask. $\mu_{p,1}^{\min}$ is the number of incoming photons that results in a SNR of 0 dB. $\mu_{p,1}^{\max}$ defines the number of incoming photons where the sensor starts to saturate. The dynamic range of the unmodified sensor is then given by $\mathrm{DR}_\mathrm{unmod} = \mu_{\mathrm{p},1}^{\max}/\mu_{\mathrm{p},1}^{\min}$.

Expressed in f-stops, this leads to

$$\mathrm{DR}_\mathrm{unmod}^\mathrm{stops} = \log_2 \mu_{\mathrm{p},1}^{\max} - \log_2 \mu_{\mathrm{p},1}^{\min}$$

Since Fig. 5 uses a logarithmic scale for the horizontal axis and since the attenuation filter of the masked pixels only lets past a fraction d_3 of the incoming photons (see also Fig. 3), the SNR curve of the attenuated pixels can be derived by a shift of the left-hand curve. The overall

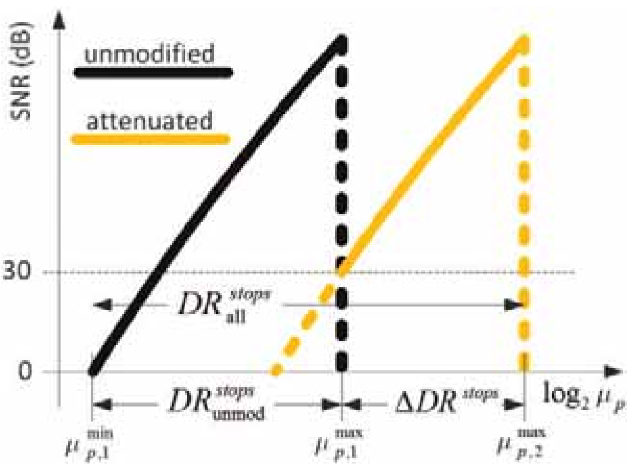


Figure 5 *Schematic SNR curve*

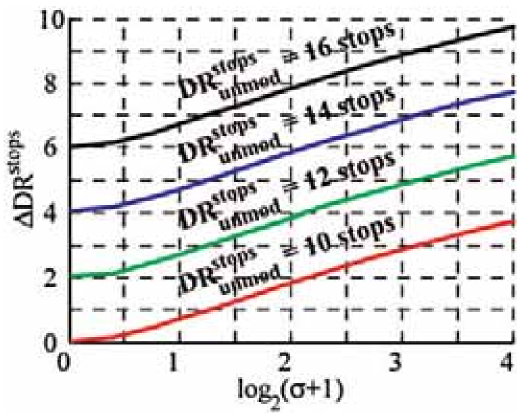


Figure 6 *Achievable dynamic range increment*

dynamic range is then given by $\mu^{min}_{p,1}$ and $\mu^{max}_{p,2}$. Obviously, the larger the attenuation factor d_3, the larger the shift and the bigger the increment ΔDR^{stops} of the overall dynamic range.

While this is a very desirable goal, it has to be noted that large shifts of the SNR curve also result in large SNR drops when switching from non-attenuated to attenuated pixels. As a consequence, image regions with an illumination slightly above $\mu^{max}_{p,1}$ will be noisy. For high-quality acquisition it is hence necessary to ensure that the SNR drop will not become visible. Psychophysical studies [11] indicate that this needs all SNR values to be larger than 30 dB, when $\mu_p \geq \mu^{max}_{p,1}$. In other words, the attenuation factor d_3 has to be chosen such that the two curves in Fig. 5 intersect at 30 dB. This obviously allows derivation of the corresponding dynamic range increment.

Fig. 6 illustrates the results. Each curve corresponds to the dynamic range increment ΔDR^{stops} for a given dynamic range DR^{stops}_{unmod} of the sensor without a mask. The horizontal axis is a logarithmic representation of the standard deviation σ from (1), measured in electrons and representing the various noise sources. Since the impact of the total quantum efficiency η is negligible, an arbitrary value can be chosen.

In essence, Fig. 6 shows that the larger the dynamic range of the unmodified sensor, the larger the achievable dynamic range increment. Moreover, the increase in dynamic range gets bigger when more read-out noise is present. While this, at a first glance, seems to be contradictory, it can be explained by the fact that sensors with significant read-out noise and large dynamic range need to have a huge peak SNR. Consequently, this leads to a bigger distance to the 30 dB threshold, opening up the possibility of large attenuation factors d_3. For cameras evaluated according to the EMVA Standard 1288, a typical lower bound for $\log_2(\sigma + 1)$ is 2.8–3.9. [According to EMVA 1288 [10], $\sigma^2 = \sigma_d^2 + \frac{\sigma_q^2}{K^2} = \sigma^2_{d,0} + \mu_I \cdot t_{exp} + \frac{\sigma_q^2}{K^2}$. t_{exp} is the exposure time, μ_I the temperature dependent dark current. $\sigma_q^2 = \frac{1}{12}$ is the quantisation noise, K the system gain. Neglecting the dark current μ_I, data sheets of EMVA 1288 compliant cameras [12] allow derivation of a typical value of σ in the range of 6–15 electrons. It should be noted that some cameras show even larger values and that one camera claims a very low value of $\sigma \approx 1.1$ electrons. The latter leads to $\Delta DR^{stops} \approx 2.75$ for a 12 f-stop sensor.] For a 12 f-stop sensor, this therefore, leads to a dynamic range increment of approximately 5 stops. [Please note that the 30 dB threshold is a conservative selection in that the SNR curves of several of today's cameras do not reach this threshold when considering medium ISO values and medium scene illumination [13]. Choosing lower thresholds leads to larger dynamic range increments.]

3.4 First prototype and manufacturing

We built a prototype for demonstrating the capabilities of the proposed approach. It is based on a regular off-the-shelf industrial greyscale camera. An attenuation mask has been fabricated and has been glued onto the surface of the sensor. With pixel-exact alignment, half of the pixels are without modification and the others are attenuated. The missing pixels are then reconstructed in software. The example images in Figs. 7 and. 8 were captured with our

Figure 7 *Results of the proposed non-regular HDR camera with logarithmic tone-mapping (left) and corresponding LDR image without tone-mapping (right)*

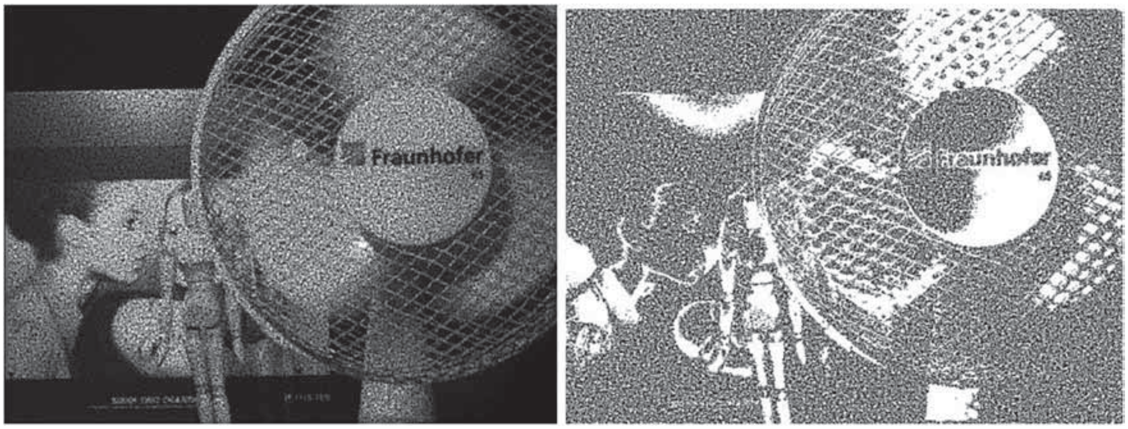


Figure 8 *Raw camera image of the proposed non-regular camera (left) and mask of valid and invalid pixels (right) marked in white and black, respectively*

prototype camera. Our prototype illustrates that the basic principle works well; however, one needs to keep in mind that gluing the mask to the sensor surface is only suitable for one-off prototypes.

For manufacturing this sensor in a large volume, better options are available: The attenuation mask can be fabricated in the metal layer of the sensor. Alternatively, one could even build pixels with different sensitivity [14]. In the long term, the proposed technique could be manufactured at no extra cost compared with a conventional sensor.

3.5 Example real-world images

The example image series was captured with our first prototype camera. The left image in Fig. 8 shows the raw camera image. In bright regions, the regular pixels are over-exposed and only the attenuated pixels deliver valid information. The resulting mask of valid (white) and invalid pixels (black) is shown in the right image. Due to the non-regular mask, this is a noise-like pattern. The final resulting reconstructed image is shown on the left of Fig. 7. Fine details are preserved and the image contains no visible artefacts. To show the gain in terms of higher dynamic range, we added an LDR image that has been captured with the same sensor but with no irregular attenuation mask. Parts of that image are over-exposed and fine details in the bright areas are lost. In comparison with Fig. 2, this shows that our camera reaches higher dynamic range with a single exposure. As designed, the rotating fan shows motion-blur but no additional distortions.

4 Conclusion

Traditional techniques for the acquisition of high dynamic range video either lead to artefacts (bracketing), or reduce recognisable details or increase sensor costs. This paper proposed a novel way for capturing high dynamic range video content. It is based on the paradigm of computational imaging using the sparsity property of natural images. To this end, a non-regular attenuation mask is placed in front of a standard sensor. By these means, its dynamic range can be increased while limiting the resolution loss. Furthermore, it avoids huge sensor costs as well as motion artefacts. Simulations with a zone plate impressively demonstrate the capability to preserve high spatial details. Furthermore, images captured with our first real demonstrator camera, proved its capability to deliver visually pleasant images. The increase in the dynamic range which can be achieved depends on the characteristics of the sensor that underlies the non-regular mask. For a 12 f-stop sensor, the final dynamic range will be between 14 and 17 f-stops, and for a 14 f-stop sensor, our approach can capture scenes with 18–21 f-stops.

Since this helps to solve the problem of the limited dynamic range of existing cameras, our next step will be the creation of a sensor design that can be reproduced on a large scale.

5 References

[1] MCCANN J., RIZZI A.: 'The art and science of HDR imaging' (Wiley-IS&T Series in Imaging Science and Technology, 2011)

[2] AKYÜZ A.O., FLEMING R., RIECKE B.E., REINHARD E., BÜLTHOFF H.H.: 'Do HDR displays support LDR content? A psychophysical evaluation'. Proceedings of ACM SIGGRAPH, 2007, vol. 26, no. 3 No. 38

[3] MANN S., PICARD R.W.: 'On being 'undigital' with digital cameras: extending dynamic range by combining different exposed pictures'. Proceedings of IS&T Annual Conference, May 1995, pp. 442–448

[4] DEBEVEC P.E., MALIK J.: 'Recovering high dynamic range radiance maps from photographs'. Proceedings of ACM

Conference on Computer Graphics and Interactive Techniques (SIGGRAPH), August 1997, pp. 369–378

[5] AGGARWAL M., AHUJA N.: 'Split aperture imaging for high dynamic range'. Proceedings of IEEE International Conference on Computer Vision (ICCV), July 2001, pp. 10–17

[6] BONNARD J., LOSCOS C., VALETTE G., NOURRIT J.M., LUCAS L.: 'High-dynamic range video acquisition with a multiview camera', *Optics, Photonics, and Digital Technologies for Multimedia Applications II*, 2012, **8436**

[7] NAYAR S., MITSUNAGA T.: 'High dynamic range imaging: Spatially varying pixel exposures'. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2000, pp. 472–479

[8] SCHÖBERL M., KEINERT J., ZIEGLER M., *ET AL*.: 'Evaluation of a high dynamic range video camera with non-regular sensor'. Proceedings of SPIE Electronic Imaging – Digital Photography IX, February 2013

[9] SCHÖBERL M., SEILER J., FOESSEL S., KAUP A.: 'Increasing imaging resolution by covering your sensor'. Proceedings of IEEE International Conference on Image Processing (ICIP), Sept 2011, pp. 1937–1940

[10] European Machine Vision Association: 'EMVA Standard 1288. Standard for Characterization of Image Sensors and Cameras. Release 3.0, November 2010

[11] XIAO F., FARRELL J., WANDEL B.: 'Psychophysical thresholds and digital camera sensitivity: the thousand photon limit', *SPIE Electron. Imag. – Digital Photography*, 2005, **5678**, pp. 75–84

[12] European Machine Vision Association: 'EMVA Standard 1288 Compliant Products', May 2013 http://www.emva.org/cms/index.php?idcatart=86

[13] http://www.dxomark.com

[14] SCHÖBERL M., BELZ A., NOWAK A., SEILER J., KAUP A., FOESSEL S.: 'Building a high dynamic range video sensor with spatially non-regular optical filtering'. Proceedings of SPIE Optics + Photonics – Applications of Digital Image Processing XXXV, August 2012, vol. 84990C

Selected content from IBC2013

ISSN 2041-5923

Improved spectral efficiency by time-frequency slicing and advanced network planning

Erik Stare

Teracom AB, Sweden

Abstract: The fundamental spectral efficiency gains that are possible using time-frequency slicing (TFS) in regular hexagon-based multi-frequency networks (MFNs) has been studied for the case of roof-top reception with a directional antenna. Four ways of building MFNs have been studied: conventional frequency reuse using a single polarisation everywhere (i) or with a systematic use of two polarisations (but only one per site) in the network (ii), using different frequency reuse patterns for each TFS frequency with a single polarisation (iii) or using two polarisations (iv), as above in (ii). The results indicate a potential spectral efficiency gain of about 50–60% when TFS with method (iv) is used, when compared with a reference non-TFS MFN case with seven frequencies and single polarisation, depending on transmitter separation distance (80km/60km) and effective transmitter height (400 m/250 m).

1 Introduction

Time-frequency slicing (TFS) is a technique whereby each broadcast service is transmitted over several radio frequency (RF) channels via frequency hopping. TFS is specified for the digital video broadcasting (DVB)-T2 [1] and DVB-next generation broadcasting system to handheld (NGH) [2] standards and offers increased RF performance, thanks to increased frequency diversity, as reported by, e.g. Makni *et al.* [3] and Giménez *et al.* [4].

In digital terrestrial TV (DTT) networks, signals of different RF channels are normally received with quite different power levels, despite being broadcast with the same effective radiated power (ERP) [4]. This allows for a TFS gain, which is the difference between the average and the minimum carrier-to-noise ratio (C/N) of the set of received RF channels. The TFS gain gives an estimate of the link budget gain for the case where coverage is defined as: reception of all transmitted services. The explanation for the TFS gain is that with TFS the coverage of *all transmitted services* is the same (at a given reception point) and is dependent on the *average* C/N of the RF channels rather than on the one with the *minimum* C/N, which is the case for the classical non-TFS case.

When interference resulting from conventional multi frequency network (MFN) frequency reuse is also taken into account, the interfering signals will have similar variations across the RF frequencies. This results in a frequency-dependent carrier-to-interference ratio (C/I), which has considerable variation. Again, the channel with the worst C/I in a given location will, for non-TFS, determine the coverage and using TFS allows for a very significant gain of the interference-limited coverage.

One interesting question is whether the increased interference immunity offered by TFS could be exploited in a tighter frequency reuse, potentially increasing the spectral efficiency. The increased spectral efficiency would result from a higher total capacity being receivable in a given spectrum. Even if a tighter frequency reuse lowers the capacity per RF channel, due to the increased interference levels, the total capacity may nevertheless be increased thanks to more RF channels being available as a result of the tighter frequency reuse. Another question is whether the network or frequency planning could be optimised, taking the characteristics of TFS into account, which may further increase the spectral efficiency. As we will see below both questions will be positively answered.

2 Spectral efficiency study

A study has been conducted within Teracom, by Carlsson and Karlsson [5], with the aim of evaluating, via computer simulations, the potential spectral efficiency gains of using TFS in interference-limited MFNs, based on the modelling and application cases below.

The MFN was built with a regular pattern of identical and equidistant omni-directional transmitters in a hexagonal lattice. Each hexagonal area has a single transmitter positioned at its centre. To cover an arbitrarily large area with such an MFN, the network needs to use a certain frequency reuse factor N. This means that N frequencies (e.g. UHF channels) are used to build the network of one multiplex using a single RF channel per multiplex and site (non-TFS). With M multiplexes there are M frequencies per site and a total of $M * N$ frequencies are required to build the complete M-multiplex network. Using TFS each site uses again M frequencies, with each of these being reused in the same or similar way as in the non-TFS case.

In an MFN, the wanted transmitter will provide a certain field strength at the reception point. Similarly, all other transmitters in the network using the same frequency will provide an interference contribution. All these interferers will add and the ratio of the wanted power to the total interfering power yields a C/I at the reception point, see Fig. 1. This C/I determines the maximum theoretical (Shannon) capacity that can be transmitted to this reception point.

The normalised (b/s/Hz) Shannon channel capacity within the used channel is given by $\log_2(1 + C/I)$.

There are many possible ways to define coverage. The criterion adopted here was to require the reception of *all*

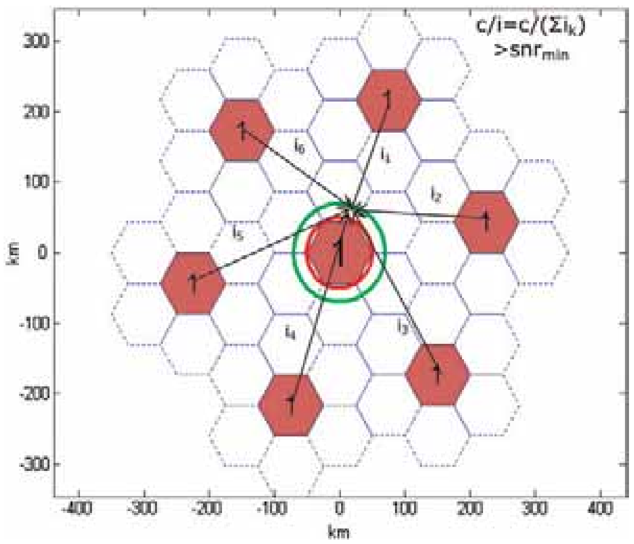


Figure 1 *Linear C/I in the reception point is given by the ratio of the wanted power c to the sum of all interferers i_k*

transmitted services (over M frequencies per site) with 95% probability [The probability refers to the reception of all services *globally* and not per individual multiplex. Having a 95% reception probability of each individual multiplex yields a probability that is lower than 95% to receive all multiplexes/services.] in the 'worst' point of the hexagon, i.e. the point with the lowest Shannon channel capacity given the actual C/I values in this point. For non-TFS this means the lowest capacity value of all the RF channels and for TFS it means the average value, which is higher. If this condition is fulfilled, then a broadcast mode could in principle be chosen (assuming an ideal system) having this capacity, which would be receivable everywhere in the hexagon (i.e. having $\geq$95% probability of receiving all services at any point within the targeted hexagon). Obviously, the capacity for non-TFS becomes lower than for TFS.

3 Propagation model

The propagation model adopted was the widely used International Telecommunication Union recommendation (ITU-R) P.1546 [6], which defines the received electrical field strength at a certain distance from a 1 kW ERP transmitter given, for example, the effective antenna height, frequency, height of receiver antenna, terrain type and percentage of time. The assumed transmitter antenna diagram is omnidirectional.

The 'effective antenna height' is the height of the antenna above terrain height-averaged between distances from 3km to 15km in the direction of the receiving antenna. Two values for effective antenna height were used: 400m and 250m, corresponding to typical values for northern and central Europe. Propagation was always over land and for 50% of time for the wanted signal and 1% of the time (i.e. increased levels) for the interfering signals. Reception was always assumed at 10 m above ground level with a directional antenna having the characteristics of ITU-R BT.419 [7], see Fig. 2. According to [7] an 'opposite' polarisation is always discriminated with 16 dB relative to the main direction of the antenna, independently of direction.

It should be noted that, since only C/I is considered in the study (i.e. interference-limited coverage), the absolute gain (frequency dependent or not) of the transmitting and

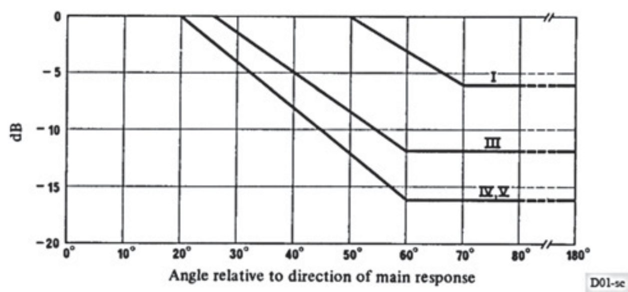


Figure 2 *Antenna discrimination according to [7]*

receiving antennas becomes irrelevant. What is very important, however, is the polar antenna *diagram* of the receiving antenna. The frequency band was always UHF band IV/V (470–790 MHz) with the used frequencies from each transmitter being evenly distributed over this band, both for non-TFS and TFS.

In addition to the deterministic propagation model given in [6], reception must also take into account fading, which needs to be statistically modelled. The total received field strength of the wanted signal as well as of all interferers is given by the deterministic field strength from [6] with additional terms for frequency-independent, but directional-dependent *shadow fading* and for frequency-dependent fading, which is unique for every transmitter.

3.1 Shadow fading

Shadow fading is modelled as log-normally distributed (i.e. having a Gaussian distribution in dB) with a standard deviation of 5.5 dB in all cases. This also follows from [6]. The shadow fading is assumed to be independent from different directions, but from a given direction all frequencies originating from a particular site are assumed to have the same shadow fading (i.e. frequency independent).

3.2 Frequency-dependent fading

Real transmitting antennas will introduce a sort of frequency-dependent fading, since a real antenna diagram typically varies significantly with frequency and direction. Also, the effects of the wave propagation (multipath) and the positioning of the antenna are frequency dependent. All these effects are modelled as an additional frequency-dependent fading of 2 dB. According to Giménez *et al.* [8], the standard deviation of the imbalance between two received RF channels with a certain frequency separation is in the range 2.5 dB–5 dB. Assuming a standard deviation of each RF frequency of 2 dB is consistent with the results presented in [8].

4 Application cases

As a reference for the performance gain, the spectral efficiency of non-TFS was used with a frequency reuse factor of $N = 7$ and with horizontal polarisation everywhere, since these are the 'classical' broadcast parameters and the ones that most resemble the actual situation for DTT in Europe. From each site six frequencies were used ($M = 6$).

To improve the spectral efficiency upon this reference case, three new technical ideas developed at Teracom were simulated in a total of four different combinations for $N = 4$, 5, 6 and 7 and were compared with the non-TFS reference case: basic TFS, TFS with multiple frequency reuse patterns (MFRP), TFS with mixed polarisation network (MPN) and TFS with MFRP *and* MPN. These are explained below. Figures within hexagons and associated colours/grey shades denote frequency.

4.1 Basic TFS

For the basic TFS case, horizontal polarisation and a single frequency reuse pattern were used for all M of the TFS frequencies. For $N = 4$, the frequency reuse pattern in Fig. 3 was used. From the centre hexagon (where the wanted transmitter is assumed to be situated) all the closest surrounding interfering transmitters appear at the same distance from the wanted transmitter. Since both the wanted signals and all interfering signals from a particular interfering site will be affected by independent frequency-dependent fading the resulting C/I will vary among the TFS frequencies.

4.2 TFS with MFRP

In this case, the TFS frequencies of a particular site have different frequency reuse patterns in the network. For $N = 4$, the three patterns shown in Fig. 4 were used.

For $M = 3$, this means that each TFS frequency gets a unique frequency reuse pattern. For $M = 6$, each pattern is used by two TFS frequencies ($M = 9$: three frequencies etc.). Thanks to the varying frequency reuse patterns, interference contributions from a particular neighbouring transmitter will only affect (for $N = 4$) *one third* of the TFS frequencies.

4.3 TFS with MPN

The receiving antenna, according to [7], has a discrimination of 16 dB (for any direction) with respect to signals of the opposite polarisation arriving within the opening angle $\pm 20°$. For signals of opposite polarisation coming from directions where there is a significant additional discrimination compared with signals having the same

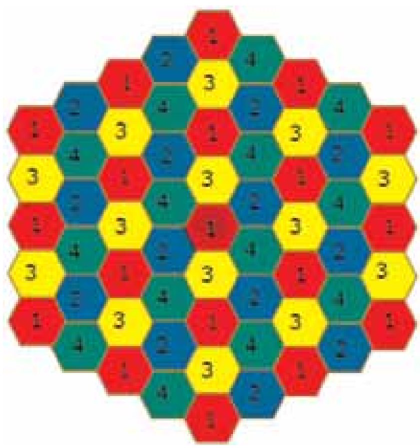


Figure 3 *Optimum single frequency reuse pattern with $N = 4$*

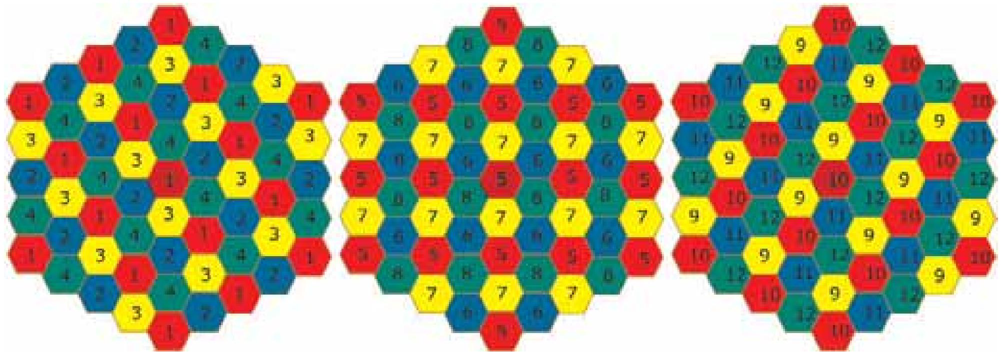


Figure 4 *Three frequency reuse patterns for N = 4 (MFRP, TFS)*
Centre hexagon uses TFS with the frequencies 1, 5 and 10, in which each has a different reuse pattern. The same principle applies to any other hexagon

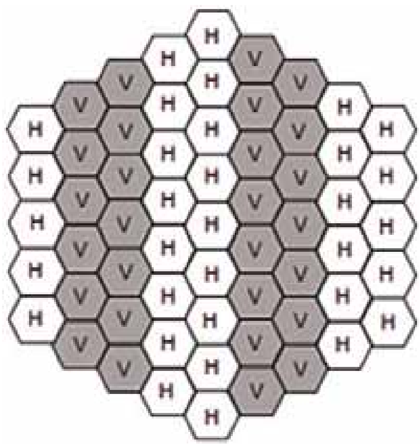


Figure 5 *Pattern of polarisations to be applied to hexagons for the MPN case*

polarisation as the wanted signal, a corresponding reduction of the interference level is achieved. It would therefore, in general, be desirable to have as many as possible of the strongest interferers using the *opposite* polarisation. The studies in [5] have shown that using the same polarisation in hexagon columns or pairs (see Fig. 5) provides a near-optimum performance together with the methods described here. In Figs. 5 and 6 'H' denotes horizontal and 'V' vertical polarisation.

4.4 TFS with MFRPs and MPN

In this case, both of the above-described additional methods (MFRP and MPN) are combined with TFS. The combination is straightforward: the polarisations in Fig. 5 are overlaid on top of the frequency reuse patterns of Fig. 4. The resulting frequency reuse patterns with polarisations are shown in Fig. 6. It should be noted that on an average only one third of the eight closest interfering transmitters use the same polarisation as the wanted transmitter, which significantly reduces the received interference thanks to the polarisation discrimination of the receiving antenna.

5 Computer simulations

To evaluate the spectral efficiency for different network configurations and other conditions, computer simulations were carried out in the form of Monte Carlo simulations. At a certain reception point, the path losses from all transmitters with significant interference contributions were calculated according to [6] as well as a large number K (typically $K = 100\,000$) of realisations from the log-normal fading distributions.

The same realisation of the shadow fading was used for all frequencies from a transmitter, but different realisations for

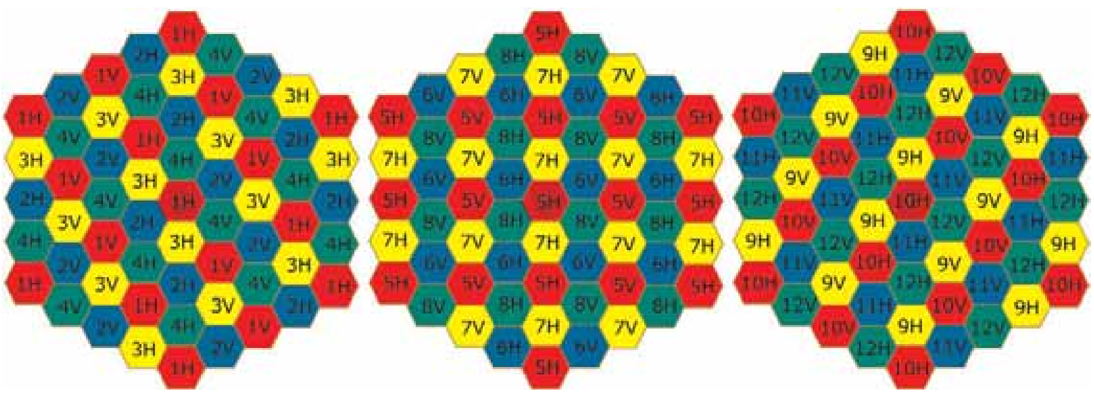


Figure 6 *Three frequency reuse patterns of Fig. 4, together with applied polarisations according to Fig. 5*

the frequency-dependent fading. In this way, it was possible to get K realisations of the wanted transmitter as well as of all interfering transmitters. For each realisation, the interference contributions were added and the C/I could be calculated as well as the corresponding spectral efficiency of the individual realisation. In this way, one could obtain K realisations of spectral efficiency for each RF frequency used.

In the TFS case, the global spectral efficiency was obtained as the average of all the M TFS frequencies. The spectral efficiency that was considered available in the reception point was the one that was available with 95% probability (or other target value) for 99% of time. The representative value for the hexagon was the spectral efficiency in the worst point of the hexagon. A special procedure was used to find the worst point quickly, so that not all the processing mentioned above would have to be performed for all points in the hexagon.

6 Results

In Tables 1 and 2, the results of the study are summarised for the case 80 km and 400 m with Table 1 showing the absolute spectral efficiency (b/s/Hz), taking into account frequency reuse, and with Table 2 showing the percentage gain in spectral efficiency compared to the case with $N = 7$ and non-TFS with six frequencies and horizontal polarisation, which is similar to real DTT implementations.

Table 1 Absolute spectral efficiency (b/s/Hz) for the 80 km/400 m

Method	$N = 4$	$N = 7$
Single frequency	1.34	1.17
Non-TFS: six frequencies	1.11	1.03
Non-TFS: six frequencies + MPN	1.37	1.20
TFS: six frequencies	1.38	1.19
TFS + MPN	1.48	1.30
TFS + MFRP	1.42	1.22
TFS + MPN + MFRP	1.54	1.32

Table 2 Spectral efficiency gain (% increase of b/s/Hz) relative to the reference case for 80 km/400 m

Method	$N = 4$, %	$N = 7$, %
Single frequency	30	14
Non-TFS: six frequencies	8	—
Non-TFS: six frequencies + MPN	33	17
TFS: six frequencies	34	16
TFS + MPN	44	26
TFS + MFRP	38	18
TFS + MPN + MFRP	50	28

'Single frequency' refers to the achievable performance with non-TFS when the coverage requirement (unrealistically) is restricted to a single multiplex, without consideration of the coverage of all other multiplexes.

With the tighter $N = 4$ reuse there is, in general, a lower received C/I compared to $N = 7$. The capacity of a *particular* RF channel is therefore on average *lower* with $N = 4$ than using $N = 7$. The reason for the higher absolute spectral efficiency with $N = 4$ in Table 1 and the higher gain in Table 2 is that the tighter frequency reuse with $N = 4$ allows for *more multiplexes* being used within a given spectrum. This gain is higher than the loss per RF channel, which results in a higher total capacity for $N = 4$.

For a 60 km transmitter distance and 250 m effective antenna height, the *absolute* spectral efficiency is, in general, lower than for 80 km and 400 m, but the percentage *gain* by the methods studied is *larger* and exceeds 60% in the best studied case (TFS + MPN + MFRP).

7 Future work

All interfering signals have so far been assumed to be perfectly correlated with respect to percentage of time, i.e. they all reach their much increased 1% percentile field strengths simultaneously. Assuming instead some degree of decorrelation between interferers, reflecting real-world observations, is likely to further increase the TFS gain. Some limited results point strongly in this direction. This is due to the fact that a TFS signal is relatively insensitive to extreme interference values of particular RF channels, as long as this is compensated for by sufficient numbers of RF channels with higher C/I. Frequency diversity will do the rest.

Similarly, very significant gains in the robustness against adjacent channel interference (from e.g. DTT, LTE or white space devices) is to be expected with TFS, as long as the effect of the interference is frequency selective. For LTE signals that are adjacent or close to DTT, more aggressive (external) receiver filtering could be used, e.g. with TFS. This is because even with a totally *nulled* DTT RF channel, caused by the filtering, the full set of services could still be received (provided there is only a small available noise margin to compensate for the increased required C/N due to the nulled RF channel).

The approach presented, TFS + MFRP + MPN, could also be applied to a single FN (SFN) case, where a group of adjacent hexagons form a local SFN cluster and these adjacent clusters use different frequencies with, e.g. frequency reuse 4. When TFS is used in such a network, the set of frequencies employed by a particular SFN hexagon group could have different frequency reuse patterns and/or polarisation patterns in a similar way as has been presented for the MFN case. It is expected that harmful effects of co-channel interference from other SFN clusters would thereby be significantly reduced.

8 Conclusions

In addition to the well-known TFS statmux and coverage gains [4], the reported study shows that in a theoretical hexagonal network, TFS and advanced network planning may provide very large improvements in fundamental spectral efficiency for interference-limited networks. By choosing the best frequency reuse factor using MFRPs and MPNs, it is possible to increase spectral efficiency by 50%, compared to a reference case with seven frequencies, for interference-limited MFNs with transmitter separation of 80 km and effective transmitter height of 400 m. Using 60 km and 250 m, the achievable spectral efficiency gain exceeds 60%.

9 Acknowledgments

The author would like to thank Håkan Carlsson and Stefan Karlsson for their excellent work, his colleague Staffan Bergsmark for valuable discussions and the IBC for publishing this paper.

10 References

[1] ETSI EN 302 755, Digital video broadcasting (DVB); frame structure channel coding and modulation for a second generation digital terrestrial broadcasting system (DVB-T2)

[2] DVB Blue Book A160 (11/12), Digital video broadcasting (DVB); next generation broadcasting system to handheld, physical layer specification (DVB-NGH)

[3] MAKNI M., ROBERT J., STARE E.: 'Performance analysis of time frequency slicing' (Fernsehseminar, 2011)

[4] GIMÉNEZ J.J., GÓMEZ-BARQUERO D., BERGSMARK S., STARE E.: 'Time frequency slicing in the next generation mobile broadcasting standard DVB-NGH'. IBC Conf. 2013, 2013

[5] CARLSSON H., KARLSSON S.: 'Network planning using time frequency slicing, polarization and directional diversity'. Teracom Internal Report, 2012

[6] ITU Recommendation ITU-R P.1546-4, 2009, Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 3000 MHz

[7] ITU Recommendation ITU-R BT.419-3, 1992, Directivity and polarization discrimination of antennas in the reception of television broadcasting

[8] GIMÉNEZ J.J., GOZÁLVEZ D., GÓMEZ-BARQUERO D., CARDONA A.: 'A statistical model of the signal strength imbalance between RF channels in a DTT network', *Electron. Lett.*, 2012, **48**, (12)

Selected content from IBC2013

ISSN 2041-5923

News service in simplified Japanese and its production support systems

H. Tanaka[1] *H. Mino*[1] *T. Kumano*[1] *S. Ochi*[2] *M. Shibata*[2]

[1]*NHK Science and Technology Research Laboratories, Japan*
[2]*NHK News Department, Japan*

Abstract: The number of foreign residents in Japan has exceeded two million but many of them have trouble understanding regular broadcasts because of the difficulty of the language used in them. We have begun researching a news service in easy Japanese to improve the accessibility of news broadcasts for these people. In this paper, we first explain the development of guidelines for simplifying regular news scripts. We then report on two support systems for rewriting the news. We consider that a team consisting of a Japanese language instructor and a news reporter is best for creating easy and precise news scripts. We developed a special rewriting editor to help them. We finally report on the launch of a public trial service called NEWS WEB EASY in April 2012. The opinions and comments sent to us revealed that deaf people and intellectually handicapped people also had high expectations for our services.

1 Introduction

There are approximately 2.08 million foreigners living in Japan today – approximately 1.6% of Japan's population. The ratio may not necessarily be high compared to that of Europe or America, but there are some areas such as Shinjuku ward, in central Tokyo whose ratio exceeds 10%. The foreign population of Japan is so diverse that it is virtually impossible to provide broadcasts in all of their native languages.

One way to solve this problem would be providing broadcasting services in simplified Japanese tailored to the language comprehension level of foreign residents. Surveys of foreign residents have shown there is a demand for broadcasts in easy Japanese [1], and we believe that there is considerable need for such services.

Our first objective was to come up with a service for converting regular news content posted on the Web, into easy Japanese content.

This involves three basic tasks: (1) figuring out the proper level of language difficulty, i.e. one that is tailored to the comprehension level of foreign residents who speak Japanese as a second language, (2) providing Web functions that further aid understanding of easy Japanese, and (3) developing a support system that translates ordinary Japanese into easy Japanese. After deciding the policies of language level and developing the Web functions and basic assistant tools, we launched a public trial service called NEWS WEB EASY in April 2012 based on machine-aided human translations of daily news articles into easy Japanese. In this paper, we will explain the above issues, go into a bit more detail about the trial service, and summarise our findings so far.

2 Basic policies of easy Japanese

The audience we have in mind is foreign residents learning Japanese as a second language who are already fairly fluent in conversational Japanese, but who now want to learn how to read news articles and the newspaper. In other words, we are focusing on foreign residents who have achieved pre-intermediate level Japanese.

We employed three basic policies in setting guidelines to rewrite regular news into an easy form. The first is the use of general tips for writing that are often mentioned in books on composition. The second is the elimination of as many news-specific hard expressions as possible. Such expressions are often problematic for students since these are not used in daily conversation. The third is the use of standard guidelines for learning Japanese as a second language. After trial rewrites of news scripts based on these policies, we discussed the problems that appeared in them with news reporters and Japanese language instructors and improved the guidelines. Below, we will show some excerpts from our guidelines that are based on the policies.

- Vocabulary

Our approach is to rewrite news articles by sticking as closely as possible to the elementary vocabulary that foreign residents learn at the initial stage of their study. We are currently using the 1600 basic words listed in the test guidelines of the 'Japanese language proficiency test' (JLPT) [2]. Note that we also use terms that are not from this list, including technical terms, proper names, and terms that frequently appear in news articles yet are difficult to reword into easier words.

- Grammar

In terms of grammar, we explain the rules for sentence length, voice, and functional expressions. News scripts in Japanese contain many long sentences. Some are longer than 150 characters. As long sentences tend to have complex syntactic structures, we rewrite them into shorter ones that are less than 50 characters in length.

We stick to active voice as much as possible. Many books on composition point out that the passive voice sounds indirect and ambiguous, and they recommend the active voice [3]. In addition, the Japanese passive voice is functionally ambiguous and is confusing to pre-intermediate-level Japanese speakers. The Japanese passive voice usually takes the endings of 'reru' and 'rareru', which coincide with expressions of possibility and honorifics. Thus, 'sawa-rareru' can mean 'be touched (passive)', 'can touch (possibility)', and 'touch' in honorific form.

Japanese news scripts have unique functional expressions related to hearsay such as 'toiu-koto-desu (it is explained that)' and 'to-shite-imasu (it has been reported).' Since they do not appear in other genres besides news, pre-intermediate-level Japanese speakers often find them difficult. One way to lower the difficulty is to change such hearsay into a normal statement. We do this when the sentence has an agent. The phrase 'keisatu-ni yore-ba 3-nin shibou-shita toiu-koto-desu (It was reported by police that three people were killed.)' has an agent 'keisastsu (police)' and the sentence is translated into 'Keisatu-ha 3-nin shibou-shita to ii-mashita (Police said that three people were killed.)' When a sentence lacks an agent, we do not change the hearsay into a normal statement.

- Curtail information

There tends to be a lot of redundancy in Web news content. This is because news copy is often written based on a script prepared to be broadcast by voice over the radio. To streamline and simplify the Japanese, we eliminated this duplication. We also cut out much supplemental information, which makes the easy Japanese version significantly shorter than the original news story.

3 Web service functions

Besides measures for simplifying Japanese itself as outlined above, we also provide a number of Web functions that aid the user in understanding the news. The following functions have been implemented in the public trial version of NEWS WEB EASY that is now available.

- Furigana (ruby) characters

Japanese is written using a combination of Chinese characters (kanji), two types of Japanese phonetic symbols (hiragana and katakana), as well as alphabetical characters (romaji) and numbers. Kanji are notoriously difficult for foreigners to master because there are so many of them and because the same characters can be read in different ways depending on the context. Foreign residents thus often find themselves unable to understand the meaning of words written in kanji. To assist them, we put very small kana characters, called furigana, above all kanji to indicate how they are pronounced. This increases the chances of foreign readers being able to understand the meaning of kanji terms even if they are unable to read them.

- Glossaries

Our basic approach is to write easy Japanese using elementary vocabulary but it is generally not possible to convert all of the difficult terms to simple vocabulary. One approach might be to add a phrase explaining such terms in a sentence but this would only lengthen the sentence and make it harder to understand. For this situation, we employ glossaries to explain difficult terminology. A glossary entry is accessed by merely positioning the cursor over the word on the NEWS WEB EASY trial screen. A popup explaining the term is then displayed. For the purposes of this trial, a dictionary for Japanese elementary school students was used to provide the glossary entries.

- Proper nouns

Proper nouns—names of people, places, organisations, and so on—are unavoidable in news articles and, of course, proper nouns are not included in any pre-existing glossaries. Different classes of proper nouns are highlighted in different colours to draw the reader's attention. The reader may not know exactly what the terms mean but, at least in this way, he or she is able to differentiate the names of people, places, and organisations.

- Text-to-speech

Some foreign residents have difficulty reading Japanese, yet are perfectly capable of understanding the text if it is read to

Figure 1 *Screen shot of the NEWS WEB EASY application*

them. NEWS WEB EASY features a text-to-synthesised voice function targeted at people who fall into this category.

Fig. 1 shows a screen shot of the NEWS WEB EASY trial, highlighting the features described above.

4 Tasks in daily services

NEWS WEB EASY serves three news stories that are manually translated from the regular news everyday except Saturday, Sunday and national holidays.

Daily work proceeds in the following order.

- Choice of the news story;
- Rewriting of the news by a news reporter and a Japanese instructor;
- Content check by a news editor;
- Adding Furigana, glossary and colouring proper nouns;
- Adding speech synthesis data;
- Finalising the web page.

The most labour-intensive part is the rewriting of the news, and we have developed two assistant tools. In the next section, we will explain the characteristics of rewriting regular news into easy Japanese and the details of the assistant tools.

5 Rewrite support system

5.1 Reciprocal translation

The rewriting from regular news into easy Japanese can be viewed as translation work but it is different from the normal language translation conducted by NHK. Let's take an example from the case of Japanese to English translation. Japanese news scripts are translated by Japanese writers (translators) and are checked and finished by an editor who is a native English speaker. Both the writer and editor have the same specialities for writing news in English, although their English competence may be different. Thanks to the shared knowledge on news translation, most of the rewriting work finishes in one pass from a writer to an editor. We thus call this type of translation 'straight human translation'.

The work for easy Japanese translation, in the meantime, requires both detailed knowledge of the regular news edition and the guidelines of easy Japanese. Since people who are well versed in the two areas are not easily to find at present, we decided to leave the translation task to a team consisting of a news reporter and a Japanese instructor who is trained in easy Japanese guidelines.

The news reporters reconstruct the news scripts by deleting redundant parts and adding background information for the audience's understanding. The Japanese instructor rewords the news story into easy Japanese basically without changing or deleting information. Since the two rewriters have different specialities, they rewrite texts from different perspectives and the rewriting goes forward and back several times between them. We call this type of translation 'reciprocal human translation'.

One of the problems of reciprocal human translation is repeated fallback of expressions. That is, the Japanese instructor changes difficult words into easier ones that often get reworded back to the original difficult one by the reporter. Some easy words are not used in regular news scripts and give an unsuitable impression to reporters. Fallback sometimes gets repeated many times, and the rewrite speed slows down. To assist with the rewrite work, we have developed a special editor [4] and a model phrase search system.

5.2 Rewrite support editor

The rewrite support editor [4] is designed to assist in reciprocal translation and has a three-column window. The left column shows the original news script, and the middle column shows the latest rewrite scripts. The right column is the work space for rewriting. The present rewriter (the Japanese instructor or news reporter) modifies the previous rewrite in the middle column while referring to the original news in the left column.

The editor shows the difficulty of words, sentences, and the whole news script in terms of numerical values and colours in real time. These help to prevent repeated fallback since they indicate the goodness of each rewrite operation in real time to give the right direction of rewriting.

The words are displayed in five colours that go with the JLPT grade [2]: blue for fourth grade, green for third grade, yellow for second grade, dark red for the first grade and light red for out of vocabulary. The rewriters pay attention to the red and yellow words, i.e. the difficult words, and change them to blue and green easy words.

Each sentence in a news script is displayed in a box with its length in characters. Sentences longer than 80 characters are shown in red, and those longer than 60 characters are shown in yellow. The rewriters shorten sentences in these colours.

The overall difficulty of a news scripts factors in three criteria: the ratio of difficult words included in the article, the average length of sentences in the article and the overall length of the article. The smaller this difficulty value, the simpler the article is. By simply checking whether the value has gone up or down after rewriting, the reporter and the Japanese instructor know whether their efforts have improved the article or made it worse.

Fig. 2 shows a screen shot of comparing various text in the rewrite support editor. The left column shows the original news scripts, the middle column shows the first rewrite by the Japanese instructor and the right column shows the second rewrite by the news reporter. The text length increased from the original script to the first rewrite. This happened because rewrites by the Japanese instructors do not delete content but add explanations of difficult expressions. The length was reduced in the second rewrite. This, of course, is done by the reporter deleting content. We can also observe that the overall difficulty fell from 5521 to 2201 and from 2201 to 1773 after each rewrite.

5.3 Model phrase search system

The rewrite support editor quickly identifies wordy or difficult passages but does not offer alternative suggestions that might improve the article. For this, we developed a database system that continually stores articles before and after rewriting. The system automatically searches for and proposes simpler phrases for a difficult phrase input by the Japanese instructor. Using the system, for example, one can easily search for an alternative phrase to substitute for the common yet difficult hearsay expression 'to-shite-imasu (it has been reported that).' This system makes it easy for rewriters to share their experiences and knowledge in the course of their work.

5.4 Translation time

We estimated the average rewriting time based on the Japanese instructors' working time. The translation time needed for the Japanese instructors is longer than that of the reporters and we consider this would give us a realistic estimate. There were two instructors and their average translation times for one article were 66.3 min. (sigma = 12.3 min.) and 73.6 min. (sigma = 13.3 min). Since the team of the Japanese instructor and reporter rewrote three news scripts at the same time and had no idle time, the figures would be close to the average time for completing an article translation.

6 Discussion

The NEWS WEB EASY trial has gone very smoothly since it was rolled out in April 2012, with three new articles released every day. So far, we have received favourable comments from foreign residents in Japan. Favourable comments also came from overseas students of Japanese. They view the NEWS WEB EASY as a learning material

Figure 2 *Comparison of original news and rewrites*

for their studies. People with handicaps also had positive comments. Those who have problems in hearing find the Furigana pronunciation guide for Kanji to be of great help for understanding people's names and new words. Furthermore, intellectually handicapped persons find the level of language suitable for their reading.

Encouraged by these comments, we are now moving on to deploying the full service and have increased the number of articles rewritten daily as of May 2013. The key issue for the full service is improving the efficiency of translation. We are now studying automatic translation technologies to be included in the rewrite support editor.

7 Conclusions

We described one of our endeavours in news services for non-native Japanese speakers, i.e. news services in easy Japanese. We first explained the basic policies for writing easy Japanese news and technologies for web services. We then elaborated the difference between regular language translation and easy Japanese translation. Easy Japanese translation is done by a team consisting of a news reporters and a Japanese instructor. We then explained two assistant tools that we have developed for the translation work and introduced our trial web service NEWS WEB EASY. Finally, we described the positive comments sent by the audience about this service.

8 References

[1] YONEKURA R.: 'Media use and information behavior of foreign residents in Japan when disasters occur'. The NHK Monthly Report on Broadcast Research Report, August 2012, in Japanese

[2] (JAPAN FOUNDATION, ASSOCIATION OF INTERNATIONAL EDUCATION: 'Japanese language proficiency test content specification' (Bonjin-sha, 2002), in Japanese

[3] STRUNK W., WHITE E.: 'The elements of style' (Longman, 2000, 4th edn.)

[4] MINO H., TANAKA H.: 'Rewrite support tool for converting new stories into easy Japanese news: targeting foreigners residing in Japan'. Annual Conference of the Institute of Image Information and Television Engineers, Japanese

ISSN 2041-5923

Interview – Jorge Rodriguez

As part of the IET and IBC's focus on young professionals working in the industry we present an interview with Jorge Rodriguez, author of the paper chosen by IBC and the IET as the best young professional contribution to IBC2013.

Tell us a bit about yourself and what you do

I received a Master's Degree in Telecommunications Engineering as well as a Bachelor's Degree in Technical Telecommunication Engineering from the University Alfonso X el Sabio in Madrid. Since the very beginning of my degree, digital television has been my area of expertise as well as one of my passions.

In 2007 I started to work in Hispasat, the Spanish satellite operator which provides services and space capacity in Europe, North of Africa and America. I have grown as a professional in the broadcast market thanks to the opportunities they gave me of participating in several projects related to future media and satellite applications, and by working closer to our customers in the Customer Engineering Department. I am also a company representative on some of the DVB technical groups including the DVB-TM (Technical Module) as well as other standardisation and industrial forums such as FAME (Future Advance Media for Europe).

What is your paper about?

The paper is about one of the hottest topics of this IBC2013: UHDTV. It describes the technological challenges faced by the industry for a successful UHDTV deployment, in particular the need for new coding and transmission systems such as the new standard HEVC (High Efficiency Video Coding) and the improvements under development in satellite broadcasting technologies for the extensions of DVB-S2.

Besides the theoretical analysis, we have included two different use cases where the gain of these new technologies is presented from a more practical approach. For example, for a typical DTH (Direct-To-Home) satellite platform, the combination of DVB-S2ext + HEVC provides an increase of 33% in the number of TV channels, including in this case 19 UHDTV services.

In order to complete the analysis, results of tests and trials that we have carried out using the Hispasat satellites are presented trying to define the main values of the key parameters for UHDTV broadcasting, including bitrate, framerate, bit depth, etc. These trials have helped us define our view about the UHDTV broadcast profile that is under discussion in the industry forums nowadays.

What interests you about this area of work?

From the very beginning of my career new TV systems and formats has been one of my favourite topics. When I started to work in Hispasat I participated in several Projects related to HDTV development and mobile TV systems. Then, when the "3D boom" arose to the market, we also worked on some 3D Projects and demo services. Taking into account all of this, UHDTV is the natural evolution in my area of expertise. I am excited with the idea of participating in the first steps of the technology, trying to define the best way to transmit this new format and collaborating in the 4K development worldwide.

How would you like to see this area develop and what do you think the challenges will be?

We are just taking the first steps of the UHDTV development, so we have to face several challenges, risks and, of course, good opportunities. In general, the main goal of the industry is try to provide the best end-to-end quality image with the lowest costs. 4K has a direct impact in the whole value chain including production, transmission, reception and display, so we have to be as efficient as possible without quality degradation. In Hispasat, as satellite operator, we focus our work in the transmission part and that is why we are collaborating in the technology developments of DVB-S2 extensions and HEVC, trying to reduce the transmission capacity needed.

Is this the first paper you have submitted to IBC, and have you been to the conference before?

I participated in the conference in 2011 talking about our experience and trials with 3DTV. I am also co-author of other papers related to hybrid networks and connected TV systems.

What will you be doing at the conference this year?

This will be a very interesting year at the IBC for us, because we have a lot of news to present. First of all, and in line with our paper work, we will announce and exhibit our UHDTV demo channel. We will also present the recently launched Amazonas-3 satellite, already in operation since March 2013, as well as the new coverage available from the newest orbital position of Hispasat Group in 36°W.

Selected content from IBC2013. This paper was voted best young professional contribution

ISSN 2041-5923

Challenges and opportunities for successful UHDTV broadcasting

J. Rodríguez A. Mourelle I. Sanz

Hispasat, Spain

Abstract: UHDTV is definitely one of the hot topics of recent years and is a key element in the future business of the audio-visual industry, from content providers to TV manufacturers. Although significant progress has been made towards its future broadcasting, there are still many challenges to address before it can become a reality. This paper describes the technological challenges faced by the industry for a successful UHDTV deployment. In particular, the need for new coding and transmission systems such as the new standard High-Efficiency Video Coding (HEVC) and the improvements under development for the extensions of DVB-S2 will be analysed. The analyses will demonstrate how these developments can be a market driver for the industry to accelerate UHDTV deployment, not only in the cinema and for sports events, but also for Direct-to-Home (DTH) satellite platforms. To fully understand the progress and the technological challenges still to be faced both from a technical and a business point of view, this paper will focus on real experiences of UHDTV transmission via satellite. It will report on the current state of the art and the quality of experience found with the experimental broadcast solutions available today.

1 Introduction

For some years now, and especially since 2012, UHDTV has been the key topic in all international tradeshows, conferences and events in the broadcast world. During the last IBC2012 and NAB2013, most players throughout the value chain: content producers, display manufacturers, network operators and encoder manufacturers, all presented their latest UHDTV products, solutions and demonstrators.

But the reality is that UHDTV is still at a very early stage and there are many challenges to surmount before it can become a reality in consumer homes.

There are currently two UHDTV profiles defined by the ITU depending on the resolution: UHDTV Level 1 or 4 K: 3840×2160 pixels and UHDTV Level 2 or 8 K: 7680×4320 pixels (also known as Super High Vision (SHV))

Associated with this new format, there are other features directly related to the technological challenges that must be addressed before a commercial UHDTV service can be conceived.

An increase in spatial resolution is directly related to the increase in the bit-rate needed for its transmission. Using current coding techniques such as MPEG-4, the bit-rate required to transmit a UHDTV channel would be around 3–4 times more than that required for an HD channel. Taking this into account, development of new coding techniques is essential for UHDTV. The industry started to consider this problem some years ago and the result is the new coding standard High-Efficiency Video Coding (HEVC) which offers a significant coding gain when compared with MPEG-4.

Apart from the bit-rate, higher resolution is also a feature of larger displays, and resolution applies temporally as well as spatially, as we shall consider throughout this paper. Frame-rate can be critical in certain types of content such as sports. Another parameter, though not as relevant as the others in UHDTV quality, is bit-depth. Some commentators have already pointed out that an increase in bit-depth would provide significant increase in the quality of UHDTV.

Along with the development of UHDTV and HEVC, important changes are taking place within the transmission standards. DVB-S2 specialists are working to update the standard to include improvements that may help the successful deployment of UHDTV.

In the following sections, a deeper description of the key technologies foreseen to play an important role in UHDTV deployment, will be provided, with a special focus on the DVB-S2 extensions work. Moreover, commercial requirements and a tentative roadmap will be presented based on the technological analysis, simulations, tests and business plan that have been created by the authors in order to clarify and define a profile for UHDTV broadcast deployment.

2 Key technologies for UHDTV development

Even though there are many technological challenges to be faced throughout the value chain before it will be possible to deploy a fully commercial UHDTV service, there are three key issues to address: better coding algorithms (e.g. HEVC) that significantly reduce the bit-rate required for a given quality, better and more efficient transmission systems (e.g. DVB-S2 extensions) and improvement of picture quality parameters for enhancing the UHDTV viewing experience.

2.1 HEVC

The development of new coding algorithms is critical for the success of certain technologies in the consumer market. The current generation of video compression technology (H.264/AVC – MPEG-4) delivers the same video quality as its predecessor (MPEG-2) at half the bit-rate and has been critical for the successful deployment of HDTV services. In an environment where demand for high quality video is continuously increasing and network infrastructures are struggling to satisfy demand, the need for even more efficient coding algorithms is more than evident. This situation is already present with HDTV and will be even more relevant when Ultra HD enters the consumer market.

The recently standardised H.265 – HEVC has been claimed to achieve a 50% average increase in compression efficiency compared with H.264/AVC, that is, provide the same video quality at half the bit-rate required by the currently best performing compression standard at a reasonable complexity increase.

The first version of the HEVC standard was published in January 2013 and most manufacturers have already presented software implementations demonstrating the possibilities it offers, with very promising results.

The details of HEVC coding are not the focus of this paper but its application, together with DVB-S2 extensions system for future satellite UHDTV platforms, are.

2.2 Higher frame-rate and bit -depth

Taking into account the increase in display size and with the aim of avoiding visual discomfort, UHDTV implementation is related to an increase in frame-rate. Typically 25–30 fps are used in current broadcast profiles, which is suitable for existing content but produces low quality of experience when used for UHDTV on large screens. Depending on the content type, at least 50fps would be recommended for UHDTV and up to 100fps in certain cases. One of the limitations of current display equipment is related to this higher frame-rate requirement; the HDMI standard defines a maximum of 25fps. Industry is currently working on the development of a new HDMI standard (HDMI 2.0) which, besides other improvements, will include these higher frame-rates.

The second important point is bit-depth for colour representation. In conventional video systems, colour is represented by 8 bits. Several industry players have recommended the implementation of 10-bit systems to avoid 'banding' effects that may occur in very high quality systems such as UHDTV. This increase of bit-depth would have a significant impact on costs, especially in UHDTV displays.

2.3 Evolution of DVB systems

Besides more efficient coding techniques, higher frame-rate and bit-depth, other improvements in communication networks are required for the transmission of UHDTV and other rich media content. Thus, evolution in transmission standards such as DVB-S2 will also be a key driver for successful deployment of UHDTV.

For this reason, since the end of 2011, DVB groups CM-BSS and TM-S2 have been working on the extension of the DVB-S2 standard under the name DVB-S2x (DVB-S2 extensions) with two main objectives: improving the spectral efficiency (bits/Hz) of the current standard and adapting it to the new uses and challenges foreseen by the satellite industry, such as unmanned vehicles, Ka band platforms or wideband transponders (WBTs).

As defined in the commercial requirements of DVB-S2x, UHDTV will without doubt, be one of the main driver applications for the deployment of this new system within television distribution services, such as Direct-To-Home (DTH) platforms, high-quality contribution and video distribution networks.

The improvement provided by DVB-S2x compared to its predecessor varies depending on the use and application. Fig. 1 represents the improvement using the most efficient

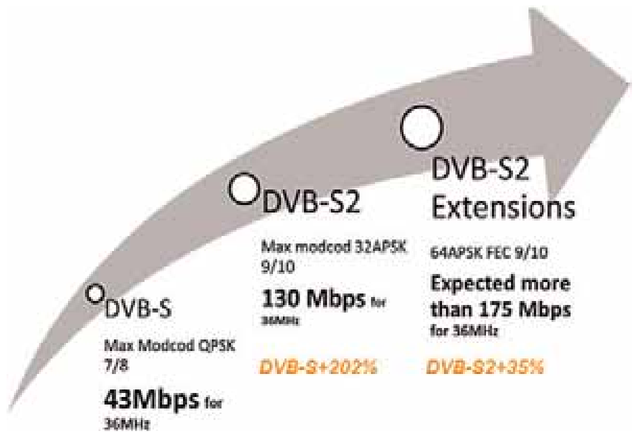


Figure 1 *DVB-S/S2 evolution*

modulation and coding (MODCOD) in each system for 36 MHz transponders. However, this improvement, quantified as approximately 35%, can be greater (near 60%) in the case of WBTs. For DTH applications, the improvement is foreseen to be within 20–35%. Even though one of the extensions was approved in 2012, the rest of system improvements are expected by the end of 2013.

The following sections present the main technologies to be incorporated that will enable the spectral efficiency improvement of the new standard.

2.4 DVB-S2 extension for WBTs

This first system enhancement was approved independently in July 2012 and introduced in the standard as Annex M to norm EN202307 (1). This annex arose with the need to update the system for transponders with high bandwidths (around 250–500MHz). Ideally, to use the maximum capacity of a transponder, it should be operated with a single carrier so that the transponder output back-off is reduced (Fig. 2). However, the complexity of the receiver for demodulating high symbol-rates, FEC decoding and filtering of such carriers greatly limits development in this field.

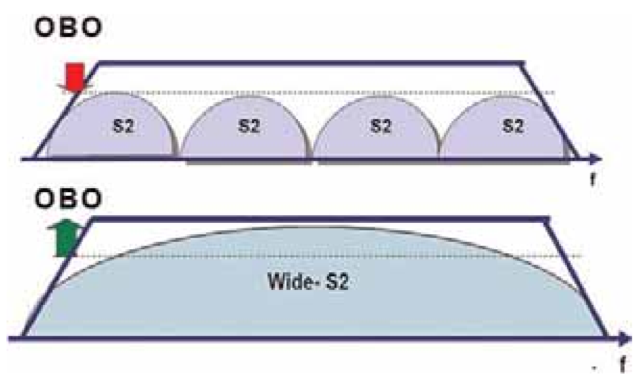


Figure 2 *Multicarrier versus single Carrier operation*

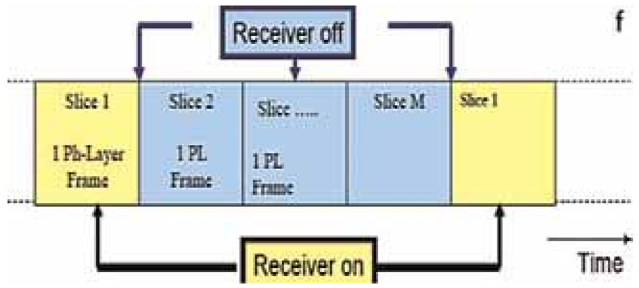


Figure 3 *Time slice concept (4)*

Because of this, DVB specialists proposed a solution based on the time-slicing concept already used in DVB-H in which several 'virtual carriers' are created within a big carrier occupying all the transponder bandwidth. Receivers demodulate the PL-Header and select only the slice of interest discarding the rest, thereby reducing significantly the receiver complexity. The time-slicing concept is shown in Fig. 3.

2.5 New modulation schemes

One of the new features that will be introduced in the new system is higher order modulation such as 64-APSK. Technology evolution has enabled the development of chipsets compatible with this modulation, providing a significant improvement in spectral efficiency. Thus, in DVB-S2, 8PSK allows a maximum spectral efficiency of approximately 2.7bps/Hz, 16-APSK permits up to 3.6 bps/Hz and 32-APSK permits 4.5bps/Hz. In DVB-S2x, however, 64-APSK will provide up to about 5.3bps/Hz. Of course, these data vary depending on the FEC implementation used.

While it is true that the use of 64-APSK in DTH platforms is difficult, it is also the case that the gain achieved with the new modulation would be useful in other applications related to UHDTV, such as contribution services and 4K distribution.

2.6 FEC granularity

Improvement in FEC granularity has no direct impact on spectral efficiency but it does impact upon it indirectly. The difference in Es/No or Eb/No between adjacent modcods in DVB-S2 is in some cases more than 1dB. This limits the option for the service provider or network operator to select the most appropriate FEC for the required quality of service, often forcing a choice of modcod which is more robust than necessary. An increase in the number of modcod options will enable a better match to be made with a particular service requirement.

2.7 Roll-off factor improvements

The roll-off factor defines how much more bandwidth the filter occupies than that of an ideal 'brick-wall' filter, whose bandwidth is the theoretically minimum Nyquist

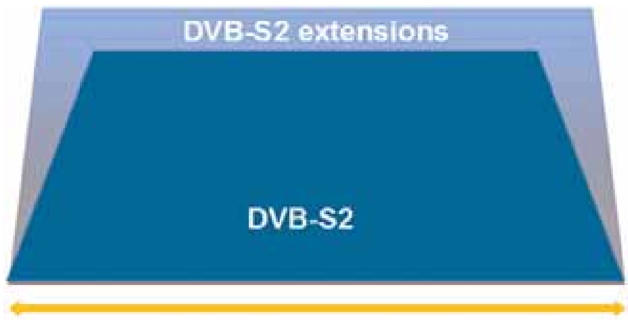


Figure 4 *Roll-off factor DVB-S2 versus DVB-S2x*

bandwidth. The roll-off factor limits the bandwidth assigned to each carrier by the formula:

$$BW = (1 + \text{Roll} - \text{Off}) \cdot \text{SymbolRate}$$

The minimum roll-off in DVB-S2 is 0.2, however, manufacturers of ground segment satellite technology (modulators/demodulators) have been improving equipment to allow more abrupt filtering, closer to the ideal. For DVB-S2 extensions, roll-off factors down to 0.05 are foreseen. This would allow an improvement of up to 18% for a typical 8-PSK carrier (Fig. 4).

2.8 DVB-S2 extensions and HEVC usage case studies for UHDTV

To fully understand the impact of DVB-S2 extensions and HEVC in the future broadcast of UHDTV services, two usage case studies over real satellite operation conditions have been developed and allow a preliminary evaluation of efficiency gain.

2.9 Ka band WBT

The first usage case is a distribution service over satellite considering Amazonas 3 Ka band Rio de Janeiro spot beam.

- Reception with 2.4m antennas with uniform distribution of terminals within spot beam.

- 450MHz transponder.

- DVB-S2: Multicarrier per transponder with 20% roll-off

- DVB-S2 extensions: Single carrier per transponder (assuming extensions for WBTs) with 5% roll-off and new modcods (estimation based on current work under development in TM-S2)

Fig. 5 shows the simulation results of ACM Maximum MODCOD in clear sky and Minimum MODCOD for target availability of 99.5% a.y. The maximum achievable bit-rate in DVB-S2 extensions for proposed use case is 1.764 Gbps, ~23% more than current with current DVB-S2.

2.10 DTH platform use case

A further very significant usage case for satellite operators and broadcasters is that of the DTH platform. For this usage case we have considered a standard satellite service over Amazonas 3 Ku band with the following considerations:

- Reception with 0.75m antennas in Brazil.

- Availability $> 99.5\%$ a.y.

- 36MHz transponders in ALC mode (single carrier).

- DTH platform with 8 transponders.

- 50% efficiency for HEVC and 40% of increment in bit-rate due to the increase from 25fps to 50fgps and 8–10 bits of bit-depth.

To provide this service the most efficient modcods for DVB-S2 and DVB-S2 ext. are summarised in Table 1.

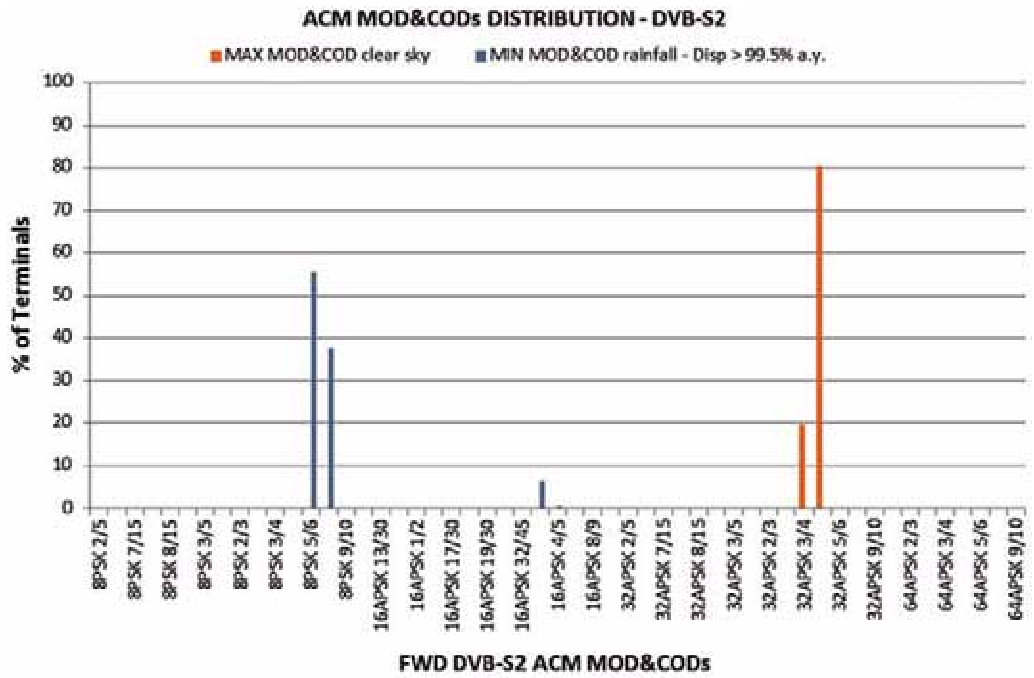
With the above data from Tables 1 and 2, Fig. 6 shows an example of a possible DTH platform design for each DVB-S2 case and both with MPEG-4 and HEVC. Only HDTV and UHDTV channels are considered. The design criteria have been: keeping the number of HDTV channels, maximising the number of UHDTV channels and include more HDTV in case of spare bit-rate.

Taking into account the results presented in Fig. 6, an increase of 33% in the number of channels could be achieved by combining DVB-S2 ext. + HEVC. This results in a DTH platform with 19 UHDTV channels as well as 37 HD channels, which is commercially reasonable for an 8-transponder platform if we compare it with a current SD + HD DVB-S/DVB-S2 platform.

3 UHDTV tests and trials

In parallel with the development of key technologies for the deployment of UHDTV, such as HEVC and DVB-S2 extensions, the last months have seen a great number of tests and trials of technology as well as demonstrators at the most important trade shows and conferences. One of the most important is the SHV format developed by NHK and demonstrated at several recent IBC shows. Also remarkable are the demo channels by Eutelsat (Quad HD format), SES (UHDTV with HEVC since April 2013) and the trials led by Abertis Telecom during MWC2013 in which a UHDTV channel was transmitted over a DVB-T2 multiplex.

Also, a group of companies lead by Hispasat performed different UHDTV trials during Q1 2013. The setup and



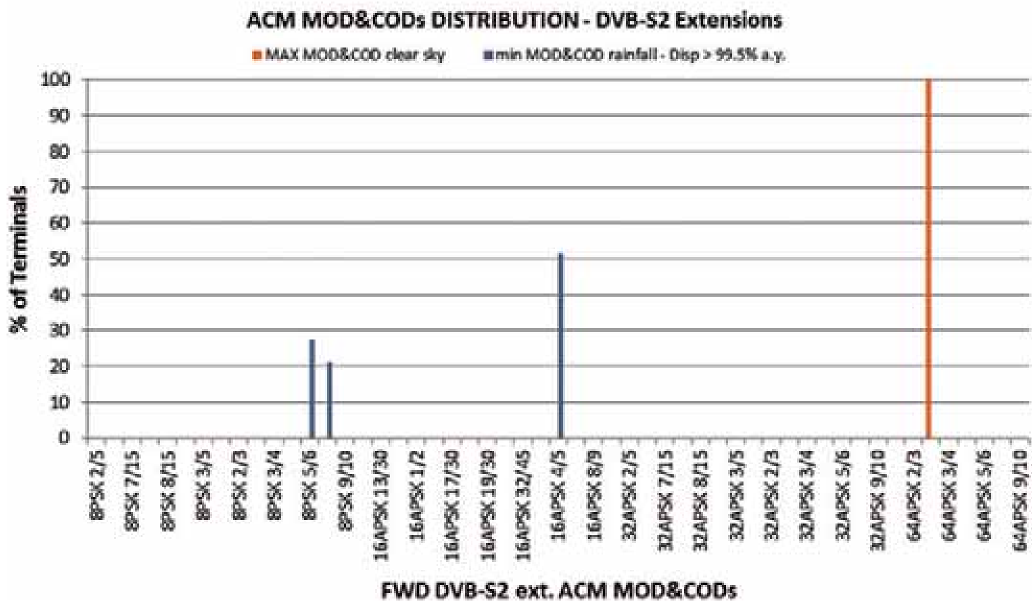


Figure 5 *MODCOD distribution using current DVB-S2 versus DVB-S2 extensions*

results of these trials are presented in the next section. The main objective was to test different transmission and coding configurations and to evaluate UHDTV signal quality and shortcomings using objective and subjective procedures. The aim is to assist commercial UHDTV transmission services to become a reality.

Table 1 Content parameters

	MPEG-4			HEVC		
	Frame-rate, fps	Bit-depth	Bit-rate, Mbps	Frame-rate, fps	Bit-depth	Bit-rate, Mbps
HDTV	25	8	7	25	8	3.5
UHDTV	25	8	35	50	10	24.5

Table 2 Performance in DVB-S2 and DVB-S2 ext. for typical DTH carrier of 36 MHz

	MODCOD	Roll-off	Bit-rate, Mbps	Threshold EbNo, dB	Spectral efficiency, bps/Hz	Clear sky margin, dB	Availability (% a.y.)
DVB-S2	8PSK 3/4	20%	65.16	4.8	1.81	3.3	99.6
DVB-S2 ext.	8PSK 3/4	5%	74.52	4.6	2.07	2.9	99.6

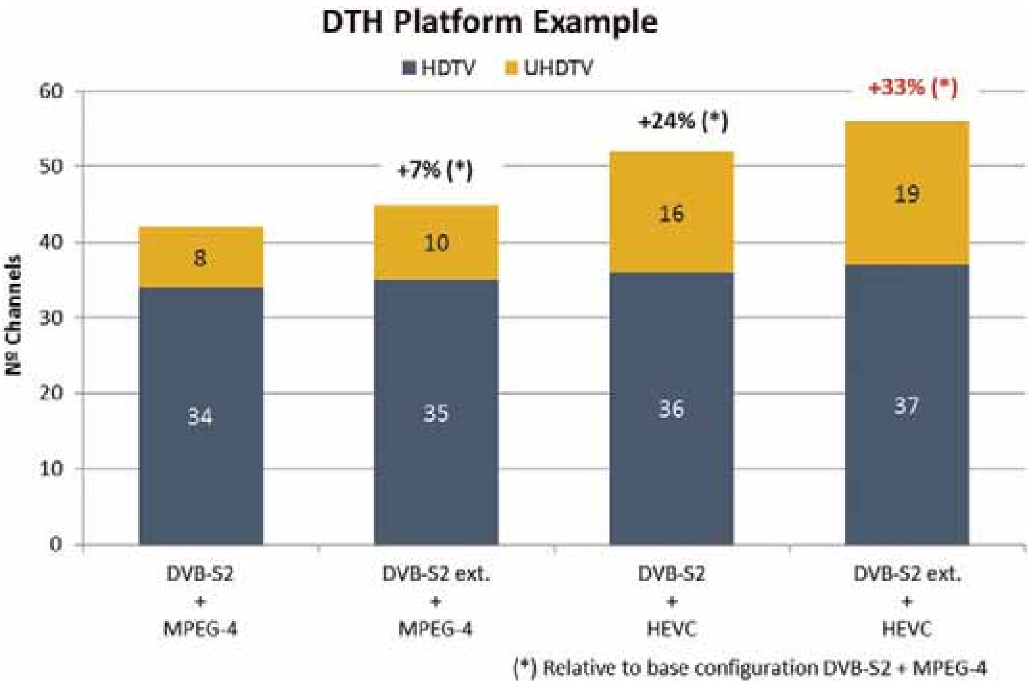


Figure 6 *DTH Platform design example*

3.1 Test architecture

The demonstrator architecture included the following elements (Fig. 7):

- *Content:* The service has a 4K loop of content available. Most of the sequences are static camera captures to avoid visual discomfort in big UHDTV displays due to the HDMI frame-rate limitation.

- *Coding:* Content was encoded using MPEG-4 with 4K profile, 25fps and 8 bits bit-depth. Trials with different bit-rates were performed to define the minimum bit-rate that would offer a high quality UHDTV service.

- *Transmission:* Different DVB-S/S2 configurations were tested.

- *Satellite:* Single 36MHz transponder operated at saturation simulating DTH conditions. Tests were performed over European coverage (using Hispasat 1E) and also South American coverage (using Hispasat 1C).

- *Receiver:* A PC with a DVB-S/S2 demodulator card was used as receiver. Content was reproduced using VLC player compatible with 4K.

- *4K display:* Two types of 4 K 84" displays were used provided by LG and Sony.

3.2 Test plan and results

During the trials three basic parameters were tested:

- *Transmission bit-rate:* Taking into account that MPEG-4 was used as the coding algorithm, bit-rates between 20 Mbps and 65 Mbps were tested. It was found that from 35 Mbps there was no perceptual quality degradation compared to high bit-rate configurations.

- *DVB-S/DVB-S2 configurations:* Different modcod configurations were tested according to the transmission bit-rate from a minimum of a DVB-S QPSK $\frac{3}{4}$ (for 35 Mbps channel) up to 8PSK $\frac{3}{4}$ for 65 Mbps channel.

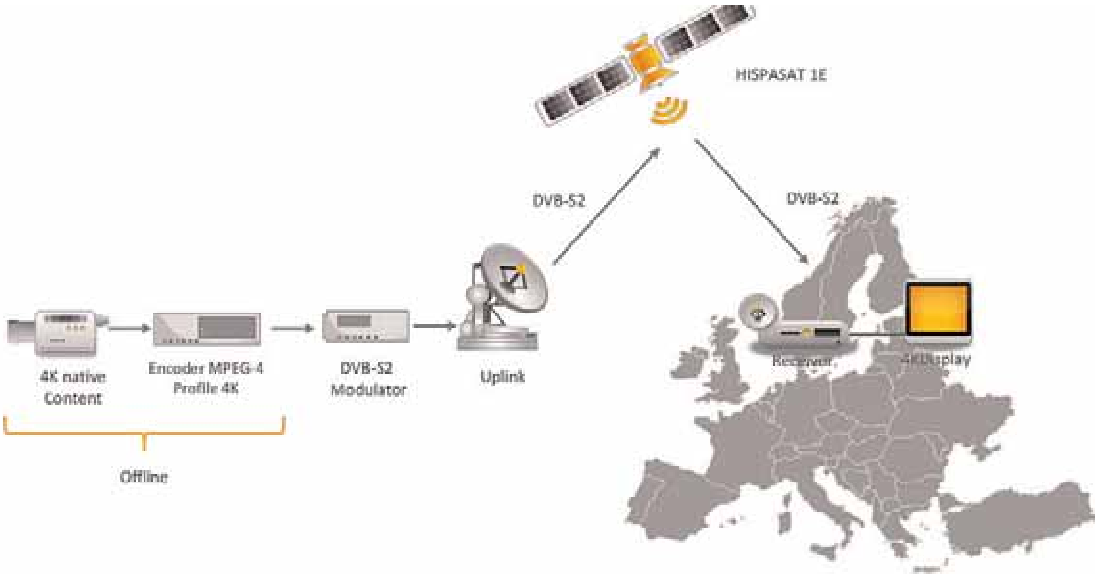


Figure 7 *4 K trials architecture*

• *Frame-rate and bit-depth:* Impact of the frame-rate and bit-depth used was verified from a subjective viewpoint with the aim of establishing the recommendations for future development of UHDTV broadcast standards. The frame-rate was the parameter where more problems were identified. The HDMI limitation of 25fps for the 4K profile significantly limited the type of content that could be transmitted with acceptable quality. With this configuration it would be very difficult to transmit commercial UHDTV services. The use of 8 bits for bit-depth did not seem to have an important impact on subjective quality perception. Some banding effects were detected even in 84" displays but without a big impact on user perception.

4 UHDTV development roadmap

Based on the technology development summarised in this paper and the forecasts from manufacturers, operators and service providers, Fig. 8 shows an estimate of the most important milestones for the development of UHDTV over satellite, mostly influenced by HEVC and DVB-S2 extensions. Without doubt, what will most influence the beginning of 4K services will be the availability of HEVC hardware and, as identified within the trials section, the availability of a higher frame-rate interface. All this leads to the conclusion that we will not see commercial UHDTV services before 2014 and high quality UHDTV services will not be available until 2015 (DVB-S2 extensions, HEVC and above 50fps).

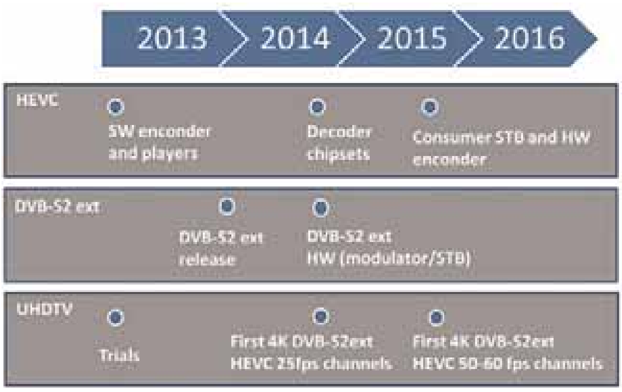


Figure 8 *UHDTV roadmap*

These prospects do not consider the development of 4K displays because 4K-compatible displays in different sizes have been available from several manufacturers since 2013. It is further, expected that prices will fall in coming months.

Another key point that goes beyond the technological developments is shown in Fig. 8, will be content availability. Currently material is scarce and mostly quite poor in quality. However, it is foreseen that in the coming months UHDTV production will increase significantly thanks to the new 4K production equipment presented in IBC2012 and NAB2013.

5 Conclusions – commercial requirements for UHDTV broadcasting

The deployment of UHDTV is clearly one of the key drivers for the growth, not only of DTH satellite platforms, but also of other broadcast platforms such as cable and DTT. In order to boost the launch of UHDTV channels it is necessary to define a 4K profile for Broadcasting. Based on the results presented in this paper, the authors propose the following:

A maximum bit-rate of 25Mbps per UHDTV channel is required in order for UHDTV to be commercially attractive for broadcasters and services providers. Otherwise, infrastructure costs will significantly limit 4 K deployment.

The availability of equipment compatible with new technologies (HEVC, DVB-S2x) will have a direct impact in the timing for UHDTV.

A minimum of 50 fps should be considered as the reference frame-rate for UHDTV broadcasting. 25–30 fps would generate a bad viewer perception of UHDTV. Even anticipation of launching 4 K 25 fps commercial services could affect UHDTV success in a negative way as it did with the launch of poor quality 3D services in the past.

Regarding the bit-depth, although no significant results were found in the trials performed it seems that 10-bit should be an option but not a mandatory parameter in the short and mid-term UHDTV future.

6 Acknowledgments

The author would like to thank the following companies involved in the UHDTV trials: RTVE, SAPEC, ERICSON, SONY and LG.

7 References

[1] ETSI EN 302 307 Digital Video Broadcasting (DVB): 'Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services', News Gathering and other broadband satellite applications

[2] WBU-ISOG 2012 Conference. Panels presentations

[3] DVB-S2 CM1250. Commercial Requirements: efficient use of ultra wide band transponders, October 2011

[4] DVB-S2 TM documentation for DVB-S2 extension. www.dvb.org

[5] EBU Draft technical report TR014. What follows HDTV? – a status report on 1080 p/50 and '4 k', April 2012

[6] Hispasat: TM-S20035: Satellite Wide Band Transponders (Hispasat contribution), November 2011

[7] ITU-R BT.1769. Parameters values for an expanded hierarchy of LSDI image formats for production and international programme exchange, May 2008

[8] MCCANN K.: 'Beyond HDTV: Implications for digital delivery' (2009)

[9] MCCANN K.: 'Progress towards high efficiency video coding' (DVB-Scene, 2012)

[10] OHM J.R., SULLIVAN G.: 'Description of High Efficiency Video Coding (HEVC)'. ISO/IEC JTC1/SC29/WG11/N11822, January 2011

ISSN 2041-5923

Introduction to *Electronics Letters*

Electronics Letters[1] is a uniquely multidisciplinary rapid publication journal with a short paper format that allows researchers to quickly disseminate their work to a wide international audience. *Electronics Letters* broad scope involves virtually all aspects of electrical and electronic technology from the materials used to create circuits, through devices and systems to the software used in a wide range of applications. The fields of research covered are relevant to many aspects of the broadcasting industry including fundamental telecommunication technologies and video and image processing.

Each year the *Electronics Letters* editorial team and the executive committee of the IET Multimedia Communications community come together to select a small number of papers from the relevant content of Electronics Letters to appear in this publication.

One of the papers chosen this year was also highlighted in the free magazine style news section included as part of each issue of Electronics Letters. The news section includes articles based on some of the best papers in each issue providing more background and insight into the work reported in the papers. A version of the associated article for this paper is also included in this publication.

We hope you will enjoy reading these papers and features as examples of our content, and if you like what you read, all our feature articles are available for free via our web pages[1].

The *Electronics Letters* editorial team

[1]www.theiet.org/eletters

Follow us on Twitter! @elecletts or scan the QR code

ISSN 2041-5923

Smart tuning

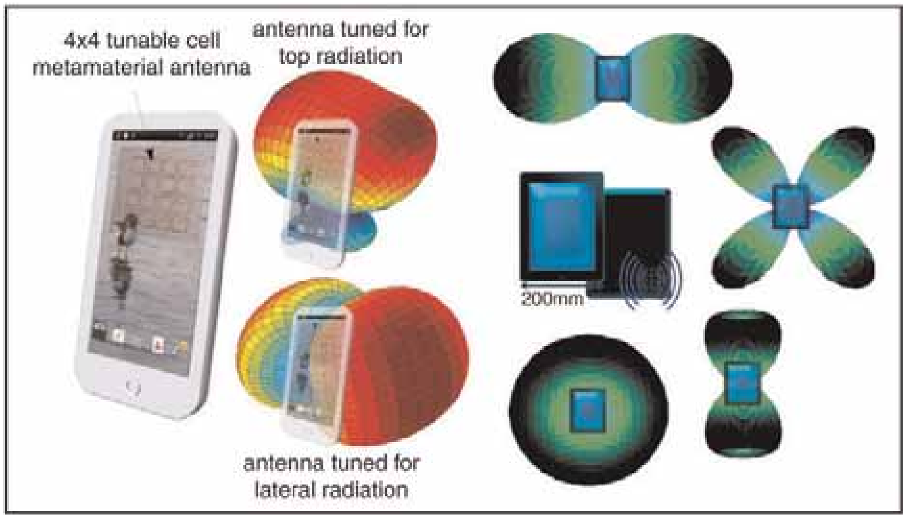
A simple multiband, multi-pattern smart antenna

A multiband smart antenna producing a range of radiation patterns without complex signal processing has been created in work from France. Suitable for microwave range use, the design could also have scaling potential for infrared and optical applications including energy harvesting.

Necessary complications?

Smart antennas use signal processing techniques to analyse and adjust the signals for each element in an array of antenna elements to achieve a specific purpose; the two main purposes being direction-of-arrival identification and beamforming.

In wireless communications they have been introduced to improve the spectrum efficiency of communication systems by exploiting a spatial domain. However, current smart antennas are often limited to a specific frequency band for a specific application. Usually, multiple frequency band systems necessitate the addition of further antennas that work at the desired frequencies. Smart antennas also often need intensive signal processing using digital signal processors which can be very complex and add computing time.



Above: *Potential applications of the metamaterial smart antenna to tune the antenna radiation patterns of a smartphone (left) and tablet device (right)*

Simple variation

In the work reported in this issue a team from the Langevin Institute of ESPCI ParisTech present a metamaterial smart antenna that can operate at multiple frequencies and generate a diverse range of radiation patterns. The planar sub-wavelength antenna uses a periodic arrangement of electromagnetic resonators and each pattern the antenna delivers corresponds to a collection of modes of this array.

"By inserting active components, these patterns can be frequency tuned from a bias voltage in order to be exploited over a wide frequency band. Moreover, using the diversity of radiation patterns, certain preferential transmission and reception directions can be selected," explained team member Camille Jouvaud.

As well as creating a multiple frequency band antenna able to provide complex radiation patterns, the proposed structure of sub-wavelength resonators is an efficient and compact alternative to traditional smart antennas. The design does not require complex technologies to implement and avoids the need for intensive signal processing by using the simple variation of a voltage bias to configure the structure.

The team attribute the success of the work to a strong understanding of cell coupling in finite and infinite arrays of sub-wavelength resonators. "The design also deals with the radiation of 'evanescent modes', i.e. collective modes with sub-wavelength spatial fluctuations. It took time for us to understand what the underlying processes are that allows such sub-wavelength modes to radiate a field in finite size structures," said Jouvaud.

The passive tuning mechanism for the array of resonators is also key to the design. The team have proposed an original method for this by coupling the resonators to a strip line where the silicon components (varactors) are embedded.

Increasing visibility

The team note that there is still some work to be done on improving the control of the antenna's directivity and bandwidth before this prototype can be turned into a commercial product, but they envision applications for it in the microwave range including wireless telecommunications, radar and detection systems. And beyond that they believe this kind of smart antenna could be applied in higher frequency ranges of the EM spectrum.

Above: *The prototype of the antenna which has a 4 × 4 grid of split ring resonators as the antenna elements. The 16 varicaps used to tune the resonant frequency can be seen on the back view (right)*

A key issue in optics is the ability to concentrate light into very small volumes. Nano antenna arrays are excellent devices for this purpose. The structure the team have proposed can be scaled to achieve a nano antenna array in the infrared and maybe at visible ranges. They have conducted FDTD simulations to prove the feasibility of the device at infrared range.

The team suggest that, to tune the resonators in this regime, Electro-optic crystals exhibiting the Pockels effect or a dielectric that shows strong Kerr effect could be used. They are currently collaborating with a group working in the infrared and optical range to explore this and Jouvaud observed that "For optical applications of metamaterials the main issue is to overcome the losses that are much larger at optical range. Nevertheless, in the infrared and visible ranges, this structure could be used as a nano antenna array for harvesting energy. For instance, this structure could be an excellent device for infrared detection or solar cells."

In related work the Langevin team have shown that it is possible to use similar media to focus energy on a sub-wavelength focal spot or to obtain a sub-wavelength resolved image of a surface only from far-field emission or measurement. More widely, the Langevin Institute team is focused on the study of propagation and control of waves in complex media, especially in media with strong heterogeneities that are especially random or periodic.

Selected content from *Electronic Letters*

ISSN 2041-5923

Adaptive metamaterial antenna using coupled tunable split-ring resonators

C. Jouvaud   J. de Rosny   A. Ourir

Institut Langevin, ESPCI-ParisTech, CNRS, UMR 7587, 1 rue Jussieu, Paris 75005, France
E-mail: camille.jouvaud@espci.fr

Abstract: A planar array of coupled split-ring resonators (SRRs) shows many resonances over a broad frequency band at which the structure efficiently emits waves. The mode splitting due to coupling offers an unprecedented way for configuring the radiation pattern of the SRR array. It is shown that specific radiation patterns that include dipolar and quadrupolar oscillation patterns can be constructed by varying the self-impedance of the resonators. Presented is an experimental demonstration based on this concept using electronically tuned SRRs. With this setup, the radiation pattern at a given frequency can be adjusted using a DC voltage command. The proposed structure is an efficient and compact alternative to traditional electronic beam-steering antennas.

1 Introduction

The development of smart antennas has gained considerable attention in recent years owing to the advance of the wireless telecommunications networks. Split-ring resonators (SRRs) have been used in the design of antennas in order to miniaturise antenna size by inductively loading the radiating element [1]. Another important application of SRRs in the field of antennas is very broad bandwidth antennas [2]. There is also significant interest in combining multiple antennas operating at different frequencies into a single aperture antenna and in dynamically steering the radiation pattern of antennas in privileged directions. SRRs have been suggested in the literature for these purposes. Such metamaterials have been used for tuning the operation frequency of antennas [3], realising optical beam-forming networks [4], as well as for achieving a phase conjugation system [5] and for achieving a tightly controlled resonance in magnetic metamaterial [6].

Recently, the concept of resonant metalens has been introduced [7]. The device is defined as a cluster of coupled resonators arranged on a subwavelength scale forming a far-field 'lens'. A realisation of this concept based on coupled SRRs is presented in [8].

In this Letter, based on this concept, we propose the design of a planar antenna made of an array of coupled SRRs. This compact antenna is working over a wide frequency domain because it shows many very close resonant modes with good matching. We investigate theoretically and numerically this behaviour and we explain how efficient radiation appropriate to each mode may be achieved. We then introduce an original way to realise a large coverage beamforming by taking advantage of this phenomenon. We made a configurable planar antenna based on this concept by adding active components to the SRRs. We demonstrate experimentally that one can dynamically tune the structure and steer the radiation pattern.

2 Structure studied

Fig. 1*a* shows a schematic view of the proposed planar antenna. It consists of a 4 by 4 periodic array of SRRs etched on 1 mm-thick Duroid dielectric substrate. The split-ring resonators are designed to show a fundamental resonant frequency around 3.5 GHz. The resonators are arranged on a square lattice with a 10 mm period. A strong interaction is expected between the resonators in this configuration. To excite this structure and use it as a planar antenna, we place a small loop antenna (5 mm diameter) just in front of one SRR as shown in Fig. 1*a*.

We have performed numerical simulations (CST Studio commercial software) to characterise the behaviour of the proposed antenna. Fig. 1*b* shows the reflection (S11)

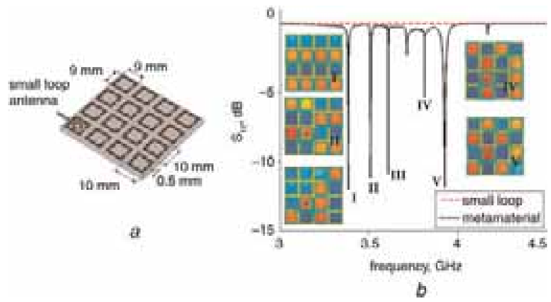


Figure 1 *Schematic view of 4 by 4 SRR array (Fig. 1a). Thickness of Duroid substrate is 1 mm. Thickness of copper layer is 30 μm. Width of SRR wire is 1 mm. Diameter of small loop antenna is 5 mm. Numerical simulation (Fig. 1b). Reflection parameter (S11) against frequency of small loop antenna. The five colour maps show amplitude distribution of magnetic normal component for five strongest modes*

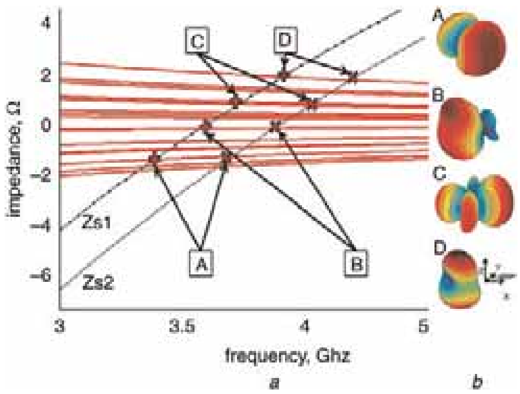


Figure 2 *Fig. 2a: red continuous line: $-\lambda_n$ where λ_n are 16 impedance eigenvalues of mutual coupling matrix. Black dash (respect. dash-dotted) line is self-impedances for cells resonating at 3.6 GHz (respect. 3.9 GHz). Fig. 2b: far-field radiation patterns for four different modes labelled A, B, C, D*

spectra of the small loop antenna in both configurations: with and without the metamaterial structure. While the injected power is almost totally reflected by the loop antenna alone, a very good matching is obtained at different frequencies around 3.5 GHz when the loop is placed near the SRR array.

3 Coupling effects in planar arrays of periodic sub-wavelength resonators

The coupling between cells induces a strong dispersion of surface waves around the SRR resonant frequency. Those resonant frequencies span over more than 500 MHz. Fig. 1*b* illustrates as well the magnetic field distribution of five significant resonances. The mode patterns can be determined and predicted with a model based on coupled dipoles [9]. In effect, the coupling between metamaterial resonators leads to the propagation of surface waves. The dispersion equation can be deduced, the Kirchhoff's laws applied on an infinite array of resonators and taking into consideration lower mutual coupling orders [8]. For a finite size array, the current in each resonator is a solution of a linear set of equations that can be concisely written in the matrix form

$$z_s\mathbf{I} + \mathbf{Z_m}\mathbf{I} = \mathbf{S} \tag{1}$$

In (1), z_s is the resonator self-impedance, Z_m is the mutual coupling matrix and S is the excitation source vector. In our case, the feeding source is the small magnetic loop that mainly couples in a non-radiative way to the SRR array. For a given excitation, the current's vector I (a complex element of the vector corresponds to a current at one cell) is found by inverting (1). Using eigenvalue decomposition of Z_m, the current vector is given by

$$\mathbf{I} = \sum \frac{\mathbf{U}_n^H \mathbf{U}_n}{\lambda_n + \mathbf{z}_s}\mathbf{S} \tag{2}$$

where λ_n and U_n are the impedance eigenvalues and the eigenvectors of the coupling matrix Z_m, respectively. The resonance occurs when $z_s + \lambda_n \sim 0$. More details can be found in [9]. It has been assumed that each SRR is well approximated by the superposition of an electric and a magnetic dipole. Then from (2) we can also compute the radiated field. Four far-field radiation patterns at four different resonant frequencies are shown in Fig. 2*b*.

We observe that even if the array is small compared to the wavelength ($\lambda/2$ by $\lambda/2$ square), the modes radiate complex radiation patterns such as dipolar and quadrupolar patterns. The higher the mode wavenumber, the higher the order of the multipole directivity pattern. Thanks to the Purcell-like effect [10], source matching occurs near resonance and efficient radiation is achieved. Indeed, we have numerically estimated an average efficiency of 80%. Hence we have built a small antenna with radiation patterns that strongly depends on the working frequency.

To utilise the benefit of the radiation pattern diversity but at one working frequency, instead of modifying the frequency, we tune the resonant frequencies of each cell by adjusting the self-impedance of the resonators. This principle is illustrated in Fig. 3*a*. Then at one frequency, it is possible to switch from one mode to another, and to obtain by this way different patterns.

To demonstrate experimentally this concept, we add a silicon diode with voltage controlled capacitance (varactors) parallel to each SRRs metallic gaps. The prototype is shown in Fig. 3*a*. Scattering parameter measurements were then performed by connecting a 5 mm diameter loop antenna to an Agilent 8722ES network analyser. Fig. 3*b* shows the return loss for several bias voltages. As observed in the simulation, for a given bias voltage, a large number of resonances appear. As expected, very good matching

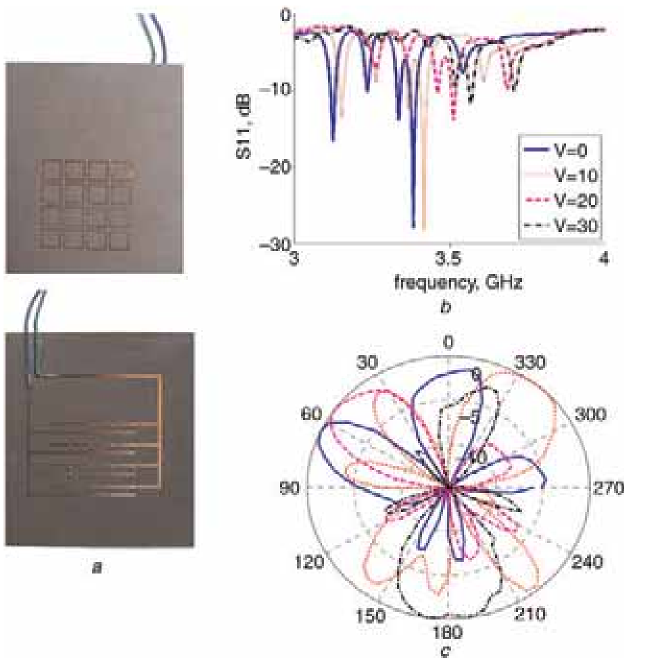


Figure 3 *Photographs of top and bottom of 4 by 4 SRR prototype (Fig. 3a). Small black parts on bottom are varicaps. Reflection parameters of 5 mm diameter magnetic loop for four different bias voltages: 0,10,20,30 V (Fig. 3b). Normalised radiation patterns at 3.31 GHz for four different bias voltages (Fig. 3c)*

(small reflection coefficient) is observed for the different resonant modes.

Now when the bias voltage increases from 0 to 30 V, the varicap capacity decreases from 6.5 to 0.5 nF and the resonances shift towards high frequencies. A tuning of the working frequency of about 200 MHz (6% of the bandwidth) is achieved. In Fig. 3*c*, we observe the effect of the bias voltage on the radiation pattern at 3.31 GHz. Thanks to the varicaps, completely different patterns are observed.

4 Conclusion

We have proposed a planar antenna based on an array of coupled SRRs. The designed antenna presents many very close resonant modes over a broad frequency band. We have observed good matching and a wide diversity of radiation patterns. Thanks to silicon diodes, we have experimentally tuned the resonant frequency of the SRRs and realised a configurable planar antenna. By scaling down our device, these results can be applied up to the infrared range where the metallic losses are still small enough. Moreover, by controlling individually each resonant frequency of the cells, we expect to increase the number of degrees-of-freedom of the cell array and to steer the emission.

5 Acknowledgments

We acknowledge the financial support of the French government-funded technological research organisation CEA/DAM. This work has been also supported by the French National Agency (ANR) with the grant OPTRANS (number 2010 BLAN 0124 04).

6 References

[1] ALICIA K.B., OZBAY E.: ‘Electrically small split ring resonator antennas’, *J. Appl. Phys.*, 2007, **101**, p. 083104

[2] KIM D.-O., JO N.-I., JANG H.-A., KIM C.-Y.: ‘Design of the ultrawideband antenna with quadripule-band rejection characteristics using a combination of the complementary split ring resonators’, *Prog. Electromagn. Res.*, 2011, **112**, pp. 93–107

[3] CHENG X., SENIOR D., YOON Y.K.: ‘Electrically small tunable split ring resonator antenna’. 2010 IEEE AP-S Int. Symp. Antennas and Propagation, Toronto, Canada, July 2011

[4] MEIJERINK A., ROELOFFZEN C.G.H., ZHUANG L., MARPAUNG D.A.I., HEIDEMAN R.G., BORREMAN A., VAN ETTEN W.: ‘Phased array antenna steering using a ring resonator-based optical beam forming network’. 2006 Symp. on the Communications and Vehicular Technology, Delft, The Netherlands, November, pp. 7–12

[5] KATKO A.R., GU S., BARRETT J.P., POPA B-I., SHVETS G., CUMMER S.A.: ‘Phase conjugation and negative refraction using nonlinear active metamaterials’, *Phys. Rev. Lett.*, 2010, **105**, p. 123905

[6] HAND T.H., CUMMER S.A.: ‘Controllable magnetic metamaterial using digitally addressable split ring resonators’, *IEEE Antennas Wirel. Propag. Lett.*, 2009, **8**, pp. 262–265

[7] LEMOULT F., LEROSEY G., DE ROSNY J., FINK M.: ‘Resonant metalenses for breaking the diffraction barrier’, *Phys. Rev. Lett.*, 2010, **104**, p. 20390

[8] OURIR A., LEROSEY G., LEMOULT F., FINK M., ROSNY J. DE: ‘Far field subwavelength imaging of magnetic patterns’, *Appl. Phys. Lett.*, 2012, **101**, p. 111102

[9] JOUVAUD C., OURIR A., DE ROSNY J.: ‘Surface waves radiation by finite arrays of magnetoelectric resonators’, *PIER*, 2012

[10] PURCELL E.: ‘Spontaneous transition probabilities in radio-frequency spectroscopy’, *Phys. Rev.*, 1946, **69**, p. 681

Selected Content from *Electronic Letters*

ISSN 2041-5923

Iterative receiver for faster-than-Nyquist broadcasting

Y.J.D. Kim   J. Bajcsy

Electrical & Computer Engineering, McGill University, 3480 University St., Montréal, QC, H3A 0E9, Canada
E-mail: yong.j.kim@mail.mcgill.ca

Abstract: Proposed is a faster-than-Nyquist (FTN) broadcast receiver architecture that is based on iterative turbo equalisation. Simulation results for this receiver show that FTN broadcasting can perform close to the capacity boundaries of the Gaussian broadcast channel. This makes FTN broadcasting a viable alternative to prior capacity achieving broadcasting techniques, e.g. superposition coding and dirty paper coding.

1 Introduction

Broadcast channels include a single transmitter and multiple receivers and occur in wireless cellular downlink transmission, radio broadcasting, satellite TV and sensor networks. It has been shown in [1] that coding in broadcast channels can in general yield higher capacity than time- or frequency-division broadcasting. Broadcast channel coding involves multiplexing messages of multiple users into one stream of channel symbols, such that the messages can be reliably recovered at the receivers. Previously known broadcast coding schemes (e.g. superposition and dirty paper coding) happen in the discrete-time domain, and the connection with continuous-time bandlimited channels is via Nyquist rate signalling and matched filtering at receivers.

Faster-than-Nyquist (FTN) signalling is a method of sending information-carrying symbols faster than the Nyquist rate of a bandlimited point-to-point continuous-time channel [2, 3]. Recently, we have proposed using FTN to achieve transmission over broadcast channels, by multiplexing signals corresponding to multiple messages using different time-offsets in the continuous-time domain, as shown in Fig. 1. In [4], we have shown that this method is capacity-wise optimal in the Gaussian broadcast channel, while it keeps the time duration and bandwidth constant. Note that in FTN broadcasting, every single channel symbol is individually and explicitly transmitted over the channel, thus removing the need for joint encoding or the definition of auxiliary signals required in previous broadcast coding schemes.

2 Continuous-time faster-than-Nyquist broadcast transmitter

Fig. 2 shows a K-user FTN broadcast transmitter with matched filter receiver front ends, originally proposed with a non-iterative receiver in [5]. Message vectors $\mathbf{m}_1$, $\mathbf{m}_2, \ldots, \mathbf{m}_K$ intended for receivers 1, 2, …, K are separately error-control encoded, interleaved, and mapped onto separate user-specific signal constellations (e.g. PAM, QAM). The resulting $N \times 1$ data-symbol vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K$ are then multiplexed into a column data-symbol vector $\mathbf{x} = vec\{[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K]^{\mathrm{T}}\}$, where $vec\{\cdot\}$ denotes matrix vectorisation and $[\cdot]^{\mathrm{T}}$ denotes matrix transpose. Symbols in $\mathbf{x}$ are subsequently FTN modulated to form a continuous-time signal:

$$\begin{aligned} x(t) &= \sum_{i=1}^{K}\sum_{n=0}^{N-1} x_i[n]s(t-nT-(i-1)T/K) \\ &= \sum_{n=0}^{KN-1} x[n]s(t-nT/K) \end{aligned} \tag{1}$$

where $x_i[n]$ denotes the nth data symbol intended for the ith receiver ($i \in [1,K]$), the Nyquist rate $1/T = 2W$ corresponds to the channel bandwidth W Hz, and $s(t)$ is a T-orthogonal unit energy modulating pulse. The available (average) transmit power P is split among K users into P_1, $P_2, \ldots, P_K$, so that $P = P_1 + \ldots + P_K$ and P_i is assigned to $x_i[n]$.

The FTN signal $x(t)$ is then broadcast to K separate receivers, where it gets perturbed by independent additive

Figure 1 *Faster-than-Nyquist signal x(t) for two-user broadcast channel; signal intended for second user is shaded for illustrative purposes*

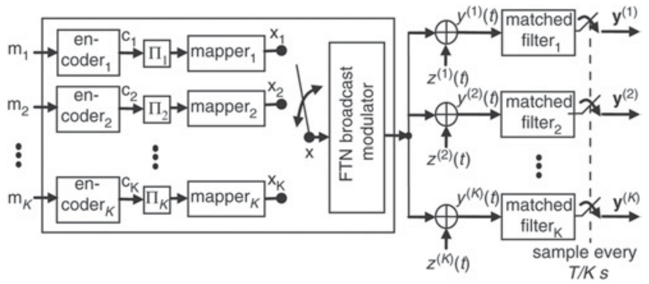


Figure 2 *System block diagram of faster-than-Nyquist broadcasting over K-user continuous-time Gaussian broadcast channel*

white Gaussian noise signals $z^{(i)}(t)$, $i \in [1, K]$, with 0 mean and two-sided power spectral densities $N_0^{(i)}/2$. At the ith receiver, $i \in [1, K]$, the noisy signal $y^{(i)}(t)$ is passed to a matched filter with an impulse response $s(-t)$ and then sampled at every T/K seconds (i.e. at the FTN signalling rate). Without any loss of generality, we will assume that the receivers are indexed according to their noise strengths $N_0^{(1)} \leq N_0^{(2)} \leq \ldots \leq N_0^{(K)}$.

3 Proposed iterative faster-than-Nyquist broadcast receiver

At the ith FTN broadcast receiver in Fig. 3, the matched filter output vector $\mathbf{y}^{(i)}$ is first passed to a maximum *a posteriori* (MAP) equaliser, along with *a priori* log-likelihood ratio (LLR) values about the data-symbols $\mathbf{x}$

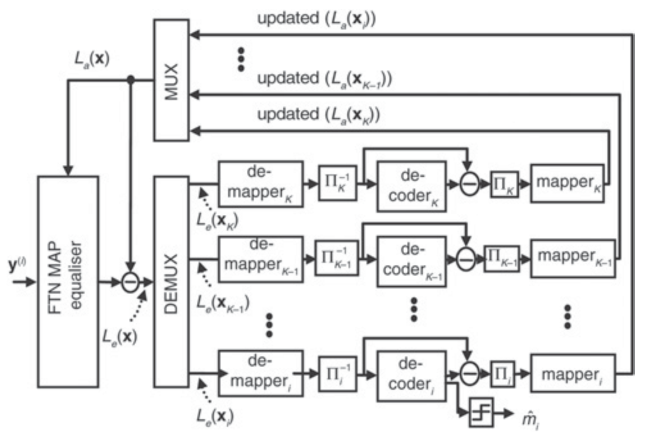


Figure 3 *Proposed ith faster-than-Nyquist broadcast receiver based on turbo equalisation principle*

(denoted by $L_a(\mathbf{x})$, which are initially set to zero). The MAP equaliser evaluates soft outputs (or reliability values) about the data symbols $\mathbf{x}$ given $\mathbf{y}^{(i)}$. The MAP equaliser is described in detail in the following Section.

From the soft outputs of the MAP equaliser, the contribution from *a priori* LLR values of $L_a(\mathbf{x})$ are subtracted to obtain an extrinsic LLR of $L_e(\mathbf{x})$ about the data-symbols in $\mathbf{x}$. Subsequently, $L_e(\mathbf{x})$ are de-multiplexed into K-user data-symbol formats, $L_e(\mathbf{x}_1)$, $L_e(\mathbf{x}_2)$, $\ldots$, $L_e(\mathbf{x}_K)$, by following the definition of $\mathbf{x}$. The ith receiver uses the degraded structure of the Gaussian broadcast channel [4] and processes only $L_e(\mathbf{x}_i)$, $L_e(\mathbf{x}_{i+1})$, $\ldots$, $L_e(\mathbf{x}_K)$, since the remaining data vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\ldots$, $\mathbf{x}_{i-1}$ cannot be properly decoded at this receiver. The extrinsic LLRs $L_e(\mathbf{x}_i)$, $L_e(\mathbf{x}_{i+1})$, $\ldots$, $L_e(\mathbf{x}_K)$ are then de-mapped into a binary bit format, which are further de-interleaved and error-control decoded. The decoder provides reliability values about the codeword bits in $\mathbf{c}_j$ and the corresponding message bits in $\mathbf{m}_j$ for $j \in [i, K]$.

In the second and subsequent iterations of the turbo equalisation, the extrinsic LLRs $L_e(\mathbf{c}_i)$, $L_e(\mathbf{c}_{i+1})$, $\ldots$, $L_e(\mathbf{c}_K)$ are re-interleaved, re-mapped, and multiplexed back together to form updated *a priori* LLR values of $L_a(\mathbf{x})$ about the data-symbols in $\mathbf{x}$. These LLR values are then fed back to the MAP equaliser for improved estimates about the data symbols in $\mathbf{x}$. The proposed iterative FTN broadcast receiver performs iterations for a prescribed number of times.

4 Faster-than-Nyquist MAP equaliser

Since FTN signalling introduces intersymbol-interference (ISI), a soft-decision MAP equaliser is needed in the receiver in Fig. 3. At the ith receiver, the nth sample of the matched filter output vector $\mathbf{y}^{(i)}$ can be written as

$$y^{(i)}[n] = \sum_{l=-L}^{L} h_l x[n-l] + z^{(i)}[n], \; n \in [0, KN-1] \quad (2)$$

where L determines the memory length of the FTN-induced ISI and the correlated Gaussian noise samples $z^{(i)}[n] = \int_{-\infty}^{\infty} z^{(i)}(t)s(t - nT/K)dt$ have zero mean and autocorrelation $E\{z^{(i)}[n]z^{(i)}[m]\} = (N_0^{(i)}/2)h_{m-n}$, $m, n \in \mathbf{Z}$. Note that L can be appropriately chosen depending on the support of pulse correlation coefficients $h_l = \int_{-\infty}^{+\infty} s(t)s(t - l \times T/K)dt$, $l \in \mathbf{Z}$.

For the MAP equaliser, an appropriate trellis diagram was constructed from the block diagram in Fig. 4. (The block diagram admits a K-tuple input $[x[n], x[n+1], \ldots, x[n+K-1]]$ and the corresponding trellis states are all possible combinations of L consecutive binary data symbols in $x[n]$.) The appropriate trellis branch metrics were obtained using [6] after realising that the FTN channel

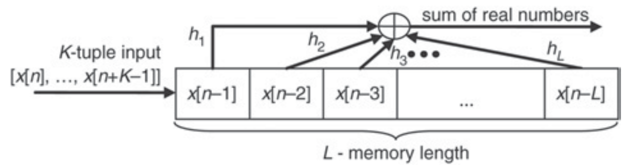


Figure 4 *Block diagram illustration of discrete-time intersymbol interference in considered FTN broadcast signalling*

model (2) is an instance of the Ungerboeck observation model. Hence, the trellis-edge branch metric between state s at the mth trellis-stage and state s' at the $(m+1)$th trellis-stage ($m \in [0, N-1]$) is given by

$$\gamma_m(s, s') = \prod_{j=Km}^{Km+(K-1)} p(x[j]) \times \exp\left(\frac{x[j]}{N_0/2}\left(y^{(i)}[j] - \frac{1}{2}x[j]h_0 - \sum_{l=1}^{L} x[j-l]h_l\right)\right)$$

Using the trellis structure and the above branch metric, the MAP equaliser was implemented using the BCJR algorithm.

5 Simulation results

Fig. 5 shows the simulated performances of the proposed FTN broadcast architecture from Figs. 2 and 3 for a two-user Gaussian broadcast channel. The modulating pulse $s(t)$ was adopted from the WCDMA standard as the square-root raised cosine with the roll-off factor $\beta = 0.22$, while time-truncation was $\pm 6T$, signalling interval $T = (1+\beta)/(2W)$ and $W = 1$ kHz. We observed that for the considered square-root raised cosine pulse and $|l| > 5$, the corresponding pulse correlations h_l were small in size (less than 0.05), hence the ISI memory length L was approximated to be 5. (It was also verified by simulations that fixing $L = 5$ indeed had only negligible performance impact.) The two encoders at the FTN broadcast transmitter were either Berrou's rate 1/2 parallel turbo codes or the rate 1/3 UMTS parallel Turbo codes. The turbo equaliser used 30 iterations with packet length $N = 2 \times 10^4$ and both users were using binary antipodal modulation. The available power $P = 1$ was split between the two user messages as $P_1 = 0.2$ and $P_2 = 0.8$.

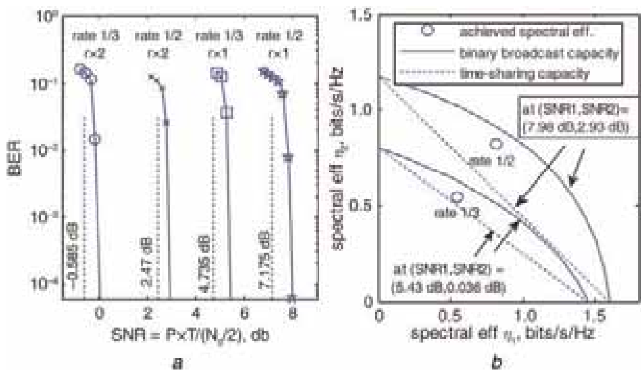


Figure 5 *Simulated performances of faster-than-Nyquist broadcast system at two receivers using either rate 1/2 or 1/3 turbo encoders*

a BER curves with respect to binary-input constrained capacities
b Achieved spectral efficiency pairs at converging SNRs

Bit error rate (BER) curves of the simulated FTN broadcast systems in Fig. 5*a* reached the target $\text{BER} = 10^{-4}$ within 1 dB from the corresponding capacities of the binary-input Gaussian broadcast channel. Fig. 5*b* shows the corresponding achieved spectral-efficiencies with respect to the capacity region of the binary-input Gaussian broadcast channel at the converging SNRs. The results demonstrate that the proposed FTN broadcast system can perform near the capacity boundary of the Gaussian broadcast channel and comparison to the dashed lines in Fig. 5*b* illustrates that it clearly outperforms time-sharing-based broadcasting. (Different power ratios were also successfully tested, yielding similar near-capacity BER performance.)

6 Conclusion

Using turbo equalisation, we propose an iterative receiver architecture for faster-than-Nyquist (FTN) broadcasting. Based on the presented results, the overall FTN broadcast architecture has been demonstrated to be a viable method for achieving the capacity region of continuous-time Gaussian broadcast channels.

7 References

[1] COVER T.M.: 'Broadcast channels', *IEEE Trans. Inf. Theory*, 1972, **18**, pp. 2–14

[2] RUSEK F., ANDERSON J.B.: 'Multistream faster than Nyquist signaling', *IEEE Trans. Commun.*, 2009, **57**, pp. 1329–1340

[3] LIVERIS A.D., GEORGHIADES C.N.: 'Exploiting faster-than-Nyquist signaling', *IEEE Trans. Commun.*, 2003, **51**, pp. 1502–1511

[4] KIM Y.J.D., BAJCSY J.: 'Faster than Nyquist broadcast signaling'. Proc. Bienn. Symp. on Communications, Kingston, ON, Canada, May 2012, pp. 186–189

[5] KIM Y.J.D., BAJCSY J.: 'An architecture for faster than Nyquist turbo broadcasting'. Proc. Int. Symp. Turbo Codes and Iterative Information Processing, Gothenburg, Sweden, August 2012, pp. 170–174

[6] COLAVOLPE G., BARBIERI A.: 'On MAP symbol detection for ISI channels using Ungerboeck observation model', *IEEE Commun. Lett.*, 2005, **9**, pp. 720–722

Selected Content from *Electronic Letters*

ISSN 2041-5923

Approach for time-scale modification of speech based on TCNMF

Haijia Wu   Xiongwei Zhang   Jianjun Huang   Weiwei Chen
PLA University of Science and Technology, Nanjing 210007, People's Republic of China
E-mail: wu_haijia@163.com

Abstract: A novel approach for time-scale modification (TSM) of speech based on temporal continuous nonnegative matrix factorisation (TCNMF) is presented. First, the magnitude spectrum of the speech is factorised to the nonnegative space and the time-varying gains, and then the TSM problem is transformed into an interpolation problem of the time-varying gains, which leads to a better performance over the traditional methods based on waveform overlap-add. The superiority of the proposed approach is confirmed by the comparative tests against the traditional methods, including OLA, SOLA, WSOLA, and PSOLA.

1 Introduction

The technology of time-scale modification (TSM) of speech can adjust the speed of a speech while keeping its perceptual features, including the pitch period, the formant structure, and so on. So it sounds like the speaker changes the speed of the speech initiatively.

Early in 1984, Griffin and Lim proposed a method called OLA [1], which divides the speech into a series of overlap-added segments by a window function and through adjusting the length of the overlap parts, the time-scale of the speech can be compressed or expanded. But the defect of this method is that the phases of the processed speech are discontinuous. To overcome this defect, Roucos and Wilgus proposed a method called SOLA [2], and Verhelst and Roelands proposed a method called WSOLA [3]. These two methods introduce an offset to correct the discontinuous phase. However, the voiced speech exhibits periodical character, and the former methods will destroy the pitch structure of the speech during their processing. This will introduce metalling sounds into the processed speech. Then, Moulines *et al.* proposed a method called TDPSOLA [4]. This method operates the speech according to the unit of the pitch periods, so it can avoid destroying the pitch structure of the speech. So it depends on accurate pitch marks, and detecting the accurate pitch marks is a challenging task.

To solve the problems existing in the traditional methods, in this Letter we propose a new TSM method based on temporal continuous nonnegative matrix factorisation (TCNMF) [5]. TCNMF is based on NMF, and by adding the temporal continuity constraint to NMF [6], the time-varying gains from TCNMF show better continuity in the time dimension. When it is used to analyse the magnitude spectrum of speech signals, such continuity is good for keeping the perceptual features of the original speech. To the best of our knowledge, TCNMF has not been previously used in TSM.

In the proposed scheme, the magnitude spectra of the original speech is factorised into a basis matrix and a coding matrix by TCNMF, and the TSM problem is transformed into a linear interpolation of the coding matrix. The perceptual features of the original speech are contained in the basis matrix [7]. In our algorithm, the basis matrix remains unchanged, so as to maintain the perceptual features of the original speech. In this way, the pitch structure of the speech can be preserved, and it can solve the problems existing in the traditional methods.

2 New TSM algorithm

There are a total of four steps in the TCNMF-based TSM algorithm, as shown in Fig. 1. First, gain the magnitude spectrum M of original speech $y(n)$ by short-time Fourier

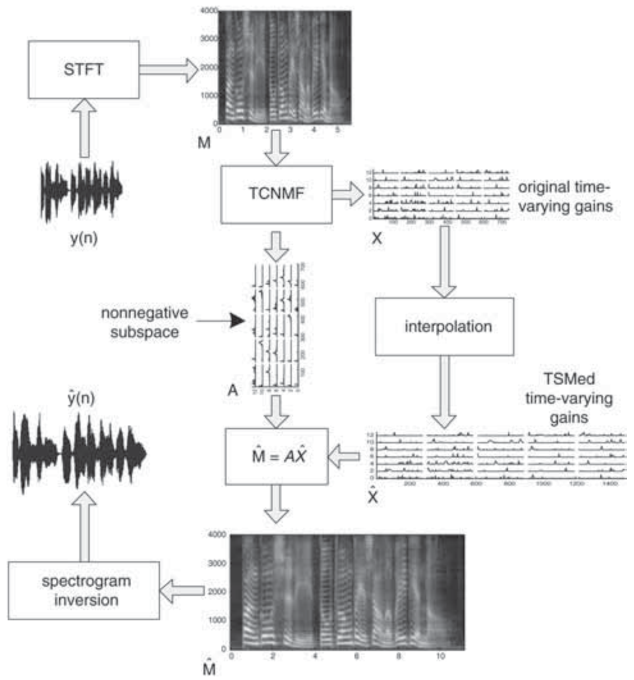


Figure 1 *Modifying time scale of voice y(n) based on TCNMF*

transform (STFT); then, factorise M to basis matrix A and coding matrix X by TCNMF; maintain the basis matrix, and gain the adjusted coding matrix $\hat{X}$ by interpolating the coding matrix according to the proportion of the time-scale modification; finally, calculate the new magnitude spectrum $\hat{M}$ by multiply A and $\hat{X}$, and recover the time-scale modified waveform $\hat{y}(n)$ from $\hat{M}$ through a spectrogram inversion algorithm. The four steps are introduced in detail as follows.

3 Magnitude spectrum representation of speech

The speech signal $y(n)$ is transformed into the frequency domain by applying a window $h(n)$ to a frame of L samples of $y(n)$, and by computing the short-time Fourier transform (STFT) of size K on the windowed data. The window is shifted by R samples before the next STFT computation. The STFT analysis results in a set of frequency domain signals that can be written as:

$$Y(k, t) = \sum_{n=0}^{L-1} y(tR + n)h(n)e^{-j2\pi kn/K}, \quad 0 \le k \le K - 1 \tag{1}$$

where $k = 0, 1, \ldots, K - 1$ is the discrete frequency index, $t = 0, 1, \ldots, T - 1$ is the frame index. Here, K is the number of frequency bins and T is the number of frames. In our implementation, we use a sampling rate of $f_s = 8000$ Hz, and $L = K = 4R = 256$. Phases are discarded, resulting in the magnitude spectrum $M(k, t) = |Y(k, t)|$. To facilitate our notation, we discard the frequency index k from $M(k, t)$, and use m_t to denote the magnitude spectrum vector $[M(0, t), M(1, t) \ldots, M(K - 1, t)]^{\mathrm{T}}$. Thus, the magnitude spectrum of a speech can be reformulated as $M = [m_0, m_1, \ldots, m_{T-1}]$.

4 Nonnegative subspace representation of magnitude spectrum

The magnitude spectrum M gained in the last step is factorised by TCNMF with rank r. As shown in 2, the result includes a nonnegative basis matrix A and a nonnegative coding matrix $\boldsymbol{X}$:

$$M = AX \quad M \in R_{+}^{K\times T}, A \in R_{+}^{K\times r}, X \in R_{+}^{r\times T} \tag{2}$$

Normally, the value of r is always smaller than K, and the relationship among r, K, and T is given in [6]: $(K + T) \times r < K \times T$.

As shown in Fig. 1, the essence of the nonnegative matrix factorisation of M is to project it to the nonnegative subspace expanded by A, and after the projection, we obtain the time-varying gains X of it in this subspace.

5 Linear interpolations of time-varying gains

Transformation of the TSM problem into linear interpolation of time-varying gains is based on the observation that the time-varying gains always change slowly over time [5]. We define the general linear interpolating function as

$$\hat{X}(i, t) = \begin{cases} \dfrac{(t\alpha - \lfloor t\alpha \rfloor)X(i, \lceil t\alpha \rceil) - (\lceil t\alpha \rceil - t\alpha)X(i, \lfloor t\alpha \rfloor)}{\lceil t\alpha \rceil - \lfloor t\alpha \rfloor}, \\ \text{s.t.} \lceil t\alpha \rceil \ne \lfloor t\alpha \rfloor \\ X(i, t\alpha), \\ \text{s.t.} \lceil t\alpha \rceil = \lfloor t\alpha \rfloor \end{cases} \tag{3}$$

where $\alpha \in R_+$ is the speed adjustment factor. If $\alpha > 1$, the speed of the speech will increase after processing, and the frame number will decrease. $i \in (0, 1, \ldots, r - 1)$, $t \in (0, 1, \ldots, T/\alpha - 1)$, r represents the row number of X, and T represents the frame number of the original speech. $\lfloor t\alpha \rfloor$ means taking the integer downwardly, and $\lceil t\alpha \rceil$ means taking the integer upwardly.

The sketch of the interpolation processing is shown in Fig. 2, where α is the speed adjustment factor, and the interpolation positions of the ith row are marked by the triangles below the line.

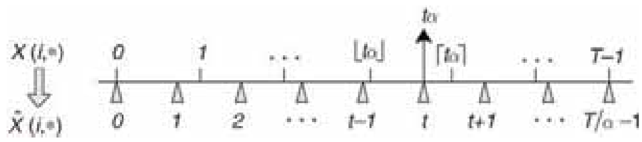


Figure 2 *Interpolation positions when speed adjustment factor is* α

6 Waveform reconstructions via spectrum inversion

After the interpolation of X, the new time-varying gains $\hat{X}$, whose time-scale has been modified, is obtained. Then, the magnitude spectrum of the time-scale modified speech can be calculated by $\hat{M} = A\hat{X}$. At last, the time-domain speech signal, whose magnitude spectrum is $\hat{M}$, can be reconstructed using the real-time spectrum inversion algorithm proposed in [8]. This algorithm can estimate the time-domain signal from the STFT magnitude spectrum without the phase information. This algorithm shows its superiority over current methods, so it is used here.

Now, the time-scale modification of the speech is complete.

7 Experiments and results

In this Section, we compare our method with the traditional methods, including WSOLA, SOLA, TDPSOLA, and OLA. As performance measures, the perceptual evaluation of speech quality (PESQ) [9] score is adopted. First, we modify the time-scale of the speeches with a speed adjustment factor α, and then re-modify the mid-speeches back to their original speeds with the speed adjustment factor $1/\alpha$, lastly the PESQ scores of the speeches after two modifications are calculated. The test speeches are 50 sentences randomly selected from the TIMIT database, and the PESQ scores are the average scores of them. From Fig. 3, we can see the performance of each method at series values of α.

When $\alpha \neq 1$, the performances of WSOLA, SOLA, and TDPSOLA are slightly lower than the performance of our method, and the performance of OLA is always far below it. The performance of TDPSOLA presents obvious fluctuation. This is because TDPSOLA depends on accurate pitch marks, and in our experiment the pitch marks of the speeches are marked automatically by machine. When $\alpha = 1$, the PESQ scores of the traditional methods used for contrast are all full marks (4.5 points), and the PESQ score of the proposed method is 4.13 points. The reason for this phenomenon is that, when $\alpha = 1$, the traditional methods do not make any modification to the speeches, but the proposed method still does TCNMF to original speech. Although the interpolation to the coding matrix is cancelled after TCNMF, the process of factorisation will make some loss to the information of original speeches.

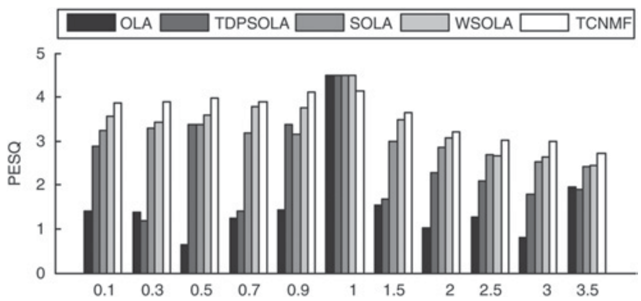


Figure 3 *PESQ scores of different algorithms at a series of* α

8 Conclusion

Compared with traditional methods, our method produces improved perceptual quality of speech. This encourages its use in many practical applications.

9 References

[1] GRIFFIN D.W., LIM J.S.: 'Signal estimation from modified short-time Fourier transforms', *IEEE Trans. Acoust. Speech Signal Process.*, 1984, **32**, (2), pp. 236–243

[2] ROUCOS S., WILGUS A.: 'High quality time-scale modification for speech'. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Tampa, FL, USA, 1985, vol. 10, pp. 493–496

[3] VERHELST W., ROELANDS M.: 'An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech'. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Minneapolis, MN, USA, 1993, pp. 554–557

[4] VALBRET H., MOULINES E., TUBACH J.P.: 'Voice transformation using PSOLA technique', *Speech Commun.*, 1992, **11**, (2–3), pp. 175–187

[5] TUOMAS V.: 'Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria', *IEEE Trans. Audio Speech Lang. Process.*, 2007, **15**, (3), pp. 1066–1074

[6] LEE D.D., SEUNG H.S.: 'Learning the parts of objects by nonnegative matrix factorization', *Nature*, 1999, **401**, (6755), pp. 788–791

[7] HUANG J., ZHANG X., ZHANG Y.: 'Recovery of lost speech segments using incremental subspace learning', *ETRI J.*, 2012, **34**, (4), pp. 255–259

[8] ZHU X., BEAUREGARD G.T., WYSE L.L.: 'Real-time signal estimation from modified short-time Fourier transform magnitude spectra', *IEEE Trans. Audio Speech Lang. Process.*, 2007, **15**, (5), pp. 1645–1653

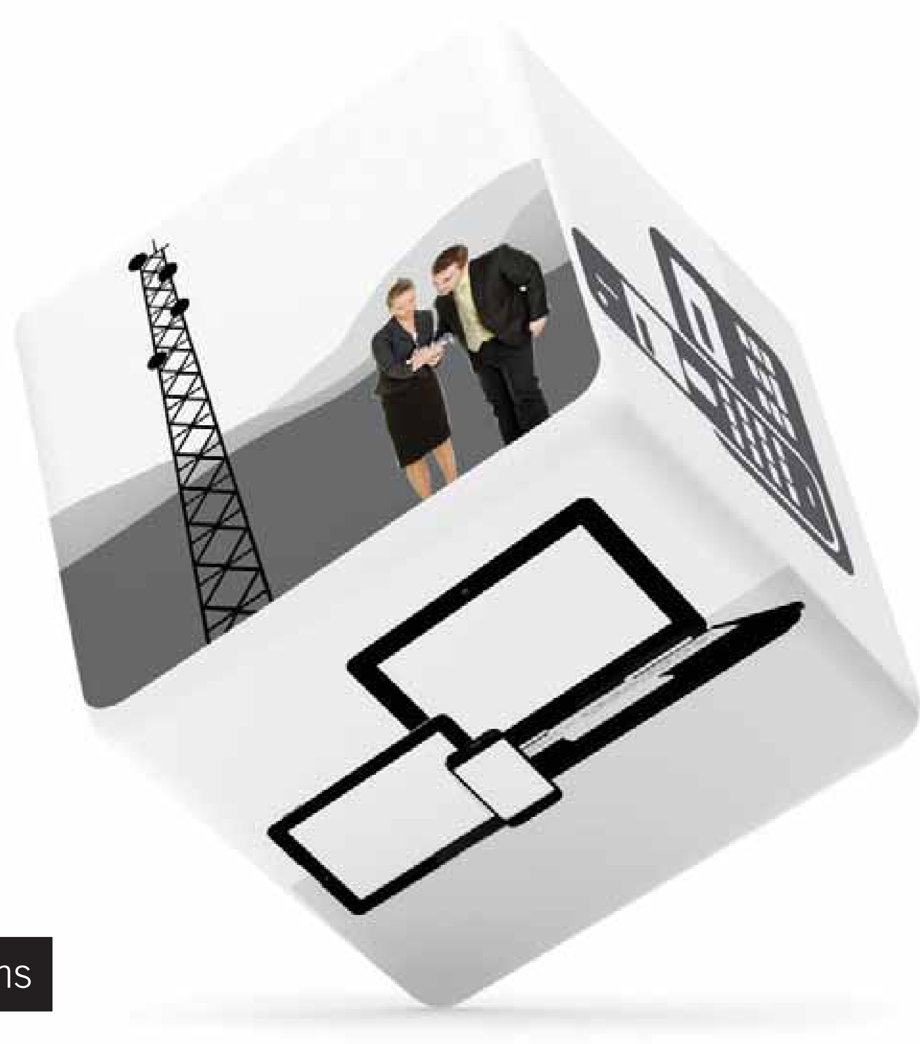
[9] Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assesesment of Narrowband Telephone Networks and Speech Codecs, ITU-T Recommendation P.862, 2001

The Institution of Engineering and Technology

Information & Communications

Connecting tomorrow, digitally

The **Information and Communications Sector** offers access to a range of products and services covering news, events, communities, professional development and support designed to help you enhance your knowledge and career.



Get involved...

www.**theiet**.org/info-comms