DRM field trials — for urban coverage planning in Spain

Unai Gil and David Guerra
University of the Basque Country

Luis del Amo and Juan Masdeu
Sociedad Española de Radiodifusión" (SER)

In 2007 the Spanish broadcaster *Sociedad Española de Radiodifusión* (SER) carried out DRM (Digital Radio Mondiale) experimental tests in close collaboration with other companies experienced in broadcasting: Vimesa Axión and the University of the Basque Country.

This article sums up the results obtained from these experimental tests. Firstly, an introductory section describes the transmission and reception infrastructures. Subsequently, the test results are summarized in three sections: *Simulcast, Monocast DRM* and *Monocast DRM Indoor Reception*.

Introduction

Various DRM field trials were carried out in the medium-wave band from July to November 2007, using an experimental network. The project focused on simultaneous broadcasting of the AM analogue and DRM digital signals using the configuration known as MCS (Multi-Channel Simulcast), as well as an evaluation of the DRM system in urban areas, considering both indoor and outdoor reception.

According to the DRM standard, there are two possible MCS configurations for ITU Region 1 [1], as shown in *Fig. 1*.

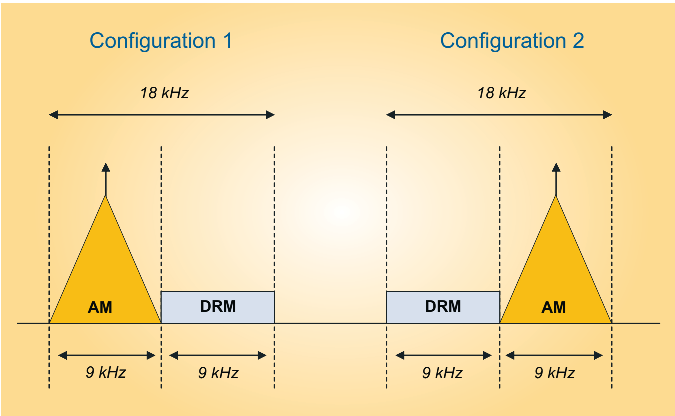


Figure 1
MCS configuration for ITU Region 1

Specifically, the objective of the tests was to find the system operating parameters for three planning scenarios:

1) ***Evaluation of the DRM-AM Simulcast System*** – to obtain the system operating parameters for planning an AM-DRM commercial service, such as the coverage area, threshold values and AM-DRM QoS by means of field measurements. The influence of the DRM signal over the AM

signal when using the Simulcast configuration would also be evaluated, as well as the optimum back-off ratio between both components of the simulcast signal: AM and DRM.

2) ***Evaluation of DRM reception in a city with dense urban areas*** – while transmitting the maximum power allowed by the experimental frequency licence, and without the power restrictions of the Simulcast configuration.

3) ***Evaluation of indoor DRM reception*** – to obtain the reception thresholds under various reception conditions inside different buildings in order to compare them with the corresponding ones for outdoor reception.

Transmission infrastructure: transmitted signals

The transmission infrastructure for the DRM tests was installed at Axion's transmitting station in Pozuelo de Alarcón, located about 9 km from downtown Madrid.

A Transradio Tram 25 transmitter with maximum peak power of 25 kW, a Fraunhofer content server, an M-Audio Delta 1010 sound system and a Transradio DRM-DMOD2 signal generator were installed, as well as a remote control system in order to facilitate measurement of different DRM and Simulcast transmission modes. The transmission system is shown in *Fig. 2*.

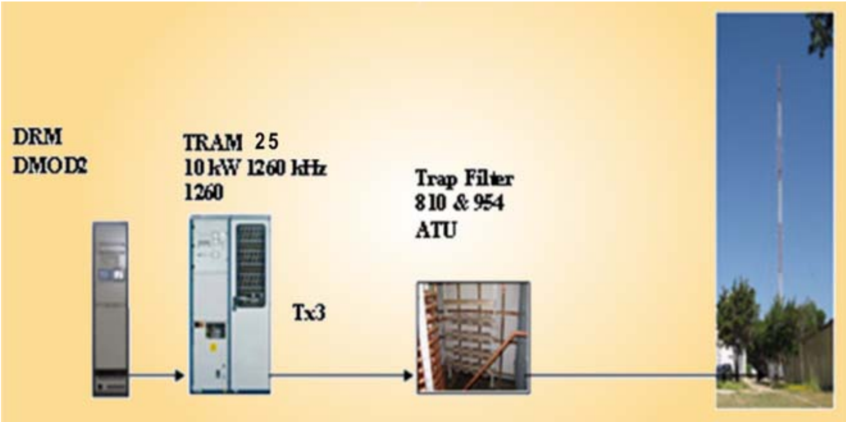


Figure 2
DRM transmission system installed in Pozuelo

The DRM-DMOD2 signal generator allows both Multi-Channel Simulcast and DRM Monocast modes. When using the Simulcast mode, the DRM signal was transmitted with a central frequency of 1251 kHz and the AM signal was centred on 1260 kHz.

The multicarrier digital signal is a potential interferer for the adjacent analogue channel. So far, the recommended back-off ratio between the AM carrier and the DRM signal has been 16 dB. This value was obtained from different laboratory tests, as well as the field trials that were carried out in Mexico City [2] and it has also been tested during the trials described in this article. In order to evaluate experimentally the possibility of decreasing this ratio, other three values have been analyzed. *Table 1* summarizes the transmitter configurations that were tested.

Table 1
MCSC transmission configurations

Power Ratio (dB)	Signal Modulation Type	DRM Power (rms) [kW]	Carrier Power (AM) [kW]	Frequency [kHz]
16	MCS	0.25	10	1251
11.8	MCS	0.5	7.5	1251
7.2	MCS	0.75	4	1251
1.5	MCS	0.9	1.25	1251

A second test phase involved measurements of DRM Monocast transmissions; i.e., only a DRM signal inside the channel. During this phase, the output transmission power was 10 kW and the digital signal was centred at 1260 kHz with a bandwidth of 9 kHz. Using this configuration, both outdoor static and mobile reception, as well as indoor reception, were evaluated.

Table 2 shows the transmission DRM mode that was selected for both the Simulcast and Monocast configurations [3]. In general, it is a robust configuration for ground-wave propagation, the only propagation mode that existed during the tests.

Table 2
DRM transmission configuration

OFDM Mode	Data Channel Constellation	Redundancy	Interleaving	Bitrate (kbit/s)	Min. SNR ITU-R BR 1615 [4] (dB)
A	64-QAM	40%	Short (0.4 s)	23.6	15.1

Reception infrastructure

The University of the Basque Country (UPV/EHU) installed a measurement system in a vehicle, specifically designed to carry out extensive measurement campaigns. *Fig. 3* shows a block diagram of the measurement system situated in the mobile unit.

The measurement system was divided into three sections: the acquisition and distribution section, the measurement equipment section and the control section. The first one consisted of a short monopole active antenna, R&S HE010, placed on top of the vehicle over a specific ground plane. The signal was distributed to various measurement devices: the R&S ESPI3 field-strength meter, the DRM professional receiver which comprised an AOR7030 front-end, a Presonus Firefox digitizer card and the Fraunhofer SW DRM Demodulator. The control system was based on a laptop

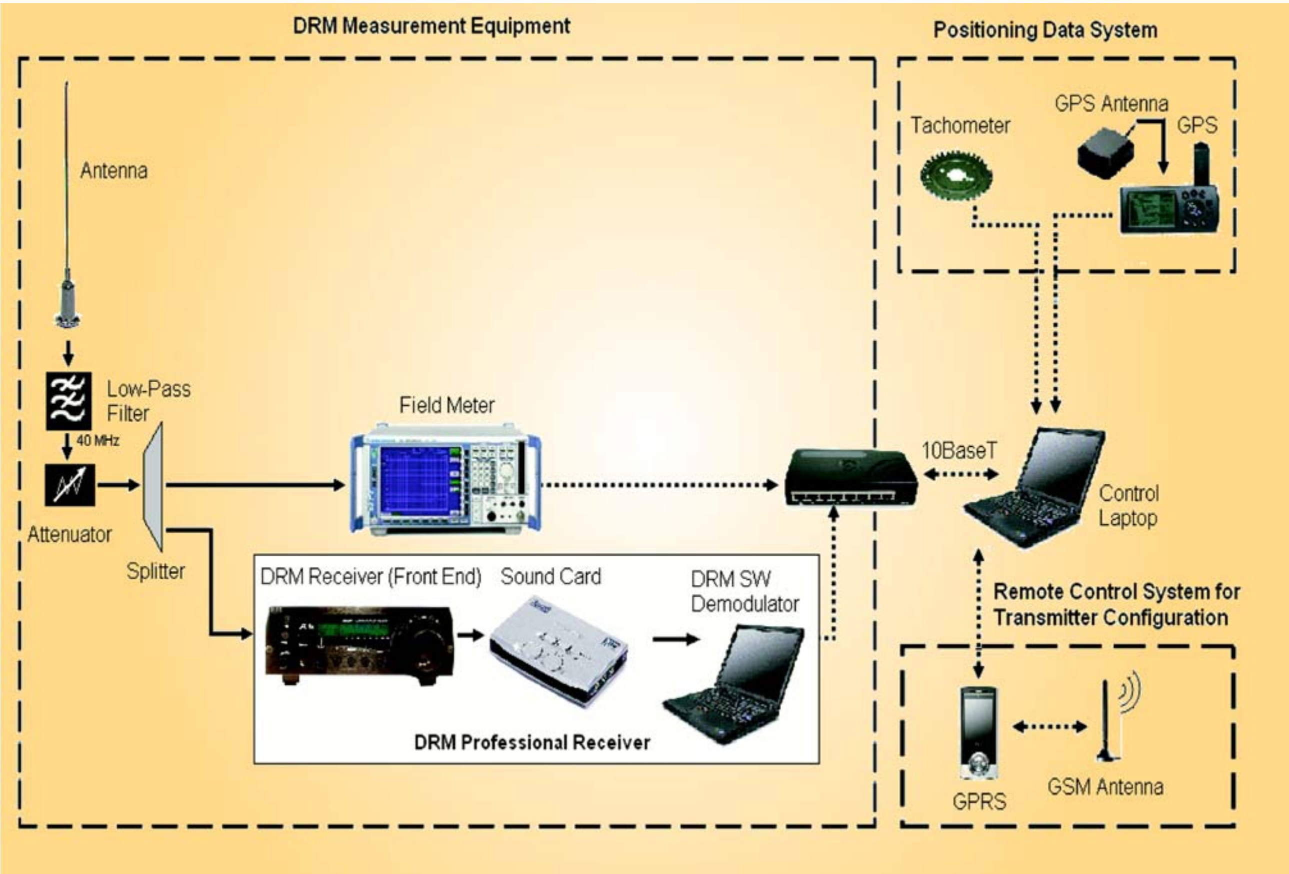


Figure 3
Block diagram of measurement system situated in the mobile unit

computer running a software control tool installed on a GNU/Linux platform in order to configure and control the rest of the equipment. In addition, a GPS receiver and a tachometer provided ancillary data such as time, position and trip data. These parameters and a set of DRM or AM signal parameters were automatically recorded by the measurement system and stored for any subsequent processing.

Regarding the indoor measurement system, a similar schema was used. The most relevant difference with respect to the outdoor system was the antenna, which was an active directional magnetic loop: a *Wellbrook Communications ALA 100*.

Simulcast tests

The so-called MCS 10001 configuration, which specifies a 16 dB of AM-over-DRM power ratio, was extensively evaluated. Overall, such a configuration was measured at more than 42 locations and along more than 210 km of roads. Radial routes or coverage routes were set up in order to assess the system coverage limits in areas further than the city of Madrid. Also, the urban routes began at the closest location to the transmitter, Pozuelo de Alarcón, but they were mainly focused on Madrid.

A second analysis – the AM/DRM mutual interference and power ratio study – involved an AM audio quality assessment using different AM receivers. In order to perform it rigorously, exhaustively and quantitatively, static reception was mandatory. Out of the 42 first locations, a subgroup of 11 reception-impaired locations was selected for measuring all four of the MCS configurations proposed for this study.

QoS has also been studied on the basis of RF measured parameters of the signal (field strength and SNR) and received audio quality parameters. The former ones were collected every 400 ms, but median values were considered for the static reception analysis. The latter ones were used to assess the availability of the analogue and digital Simulcast services.

The availability of the analogue service was assessed by the evaluation of the audio demodulated by commercial receivers. Subjective AM audio quality assessment was carried out in accordance with the audio quality degradation criteria of ITU-R Rec. BS.1284 [5]. Hence, each evaluator assessed an audio sample according to this degradation scale. On this scale, values of “4” (the degradation is perceptible but not annoying) and “5” (imperceptible degradation) were considered as “satisfactory”, meaning that correct reception was obtained.

Taking into account the previous procedure, an audio sample has been considered to show satisfactory quality, that is, to show AM service coverage, if the mean value of its evaluation scores was higher than “3.5”. On the other hand, at each location, every listener evaluated the AM service using all the proposed receivers, so that there was a mean evaluation score for each receiver.

Abbreviations

16QAM 16-state Quadrature Amplitude Modulation
64QAM 64-state Quadrature Amplitude Modulation
AM Amplitude Modulation
DRM Digital Radio Mondiale
http://www.drm.org/
ETSI European Telecommunication Standards Institute
http://pda.etsi.org/pda/queryform.asp
GPRS General Packet Radio Service
GPS Global Positioning System
GSM Global System for Mobile communications
ITU International Telecommunication Union
http://www.itu.int
ITU-R ITU - Radiocommunication Sector
http://www.itu.int/publications/sector.aspx?lang=en&sector=1
MCS Multi-Channel Simulcast
MER Modulation Error Ratio
MW Medium-Wave
QoS Quality of Service
SNR Signal-to-Noise Ratio
SW SoftWare
Tx Transmitter

In the Simulcast trials, six AM receivers were used in order to obtain AM-quality quantitative evaluations. From now on, these receivers will be referred to as A, B, C, D, E and F. Receiver A was a high-end one. Receiver B was an upper mid-range AM receiver. The rest were a representative set of commercial mid-range receivers. Each AM audio sample which was obtained from one receiver at a certain location, was evaluated by four expert listeners. Likewise, AM Monocast QoS was evaluated as a subjective audio quality reference of the AM service without the influence of the Simulcast DRM adjacent signal.

The subjective audio quality for a certain DRM service depends on the bitrate provided by the transmitted DRM configuration. However, only if the DRM signal is correctly decoded in the receiver will the intended audio quality be enjoyed by the listener. If the DRM signal is not correctly received, audio frames could be lost, thus causing audio dropouts. In this way, it is possible to obtain objective quality values for the DRM service as a function of the correctly decoded audio frames at the receiver by means of the so-called AudioQ (AQ) parameter, as follows:

$$AudioQ(\%) = \frac{Number\ of\ correctly\ decoded\ audio\ frames}{Number\ of\ transmitted\ audio\ frames}$$

Given this objective parameter, a value of 98% is found to be the minimum for static reception, below which the average listener begins to detect audible degradation of the subjective audio quality. This threshold has been taken as a reference value for satisfactorily received audio quality and it has been used in many trials within the DRM Consortium. According to the time statistical error distribution, this value is a very conservative criterion for time intervals of several minutes.

AudioQ has also been considered for DRM mobile quality assessment. However, in this case, instead of calculating it every three minutes, it has been obtained every 400 ms – the duration of one DRM frame containing 10 audio subframes. With this AudioQ figure, one erroneous audio subframe within a DRM frame was accepted as still providing satisfactory reception, thus it is a bit less restrictive than the 98% figure.

MCS 10001 results (16-dB power ratio)

The global results obtained from the static reception trials of the analogue part of the MCS 10001 configuration are shown in *Table 3*. On an area by area basis, the AM service coverage near the transmitter is rather good using the high-end receiver and acceptable for the mid-range receiver. On the other hand, reception in downtown Madrid was very good using the high-end receiver but unsatisfactory with the mid-range receiver at half of the locations (which were usually located in narrow streets). Nevertheless, the AM service degradation was not very great.

Table 3
MCS 10001 AM quality global results

AM Power (kW)	10.0
Total number of assessed locations	32
% of locations with satisfactory reception (average > 3.5). Best receiver	90.6%
% of locations with satisfactory reception (average > 3.5). Worst receiver	43.7%

A subjective evaluation of the received audio quality with two receivers simultaneously was not possible for mobile measurements. However, along those stretches where DRM measurements were also carried out, an estimated comparison between the qualities of the DRM and AM services was done using the high-end receiver. Overall, the AM reception results for the high-end receiver were very good; better than the DRM results. Following the East Radial route, starting from the transmitter and crossing the centre of Madrid, satisfactory audio quality was obtained up to distances of 13 km away from the transmitter. In the case of the South Radial, also from the transmitter but not across Madrid, the maximum distance where the AM service was received satisfactorily exceeded 80 km in rural areas.

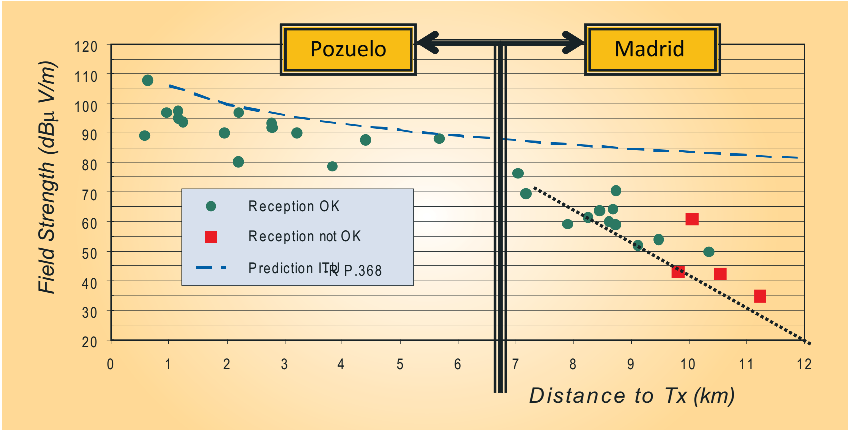


Figure 4
DRM Electric field strength marking points with correct and incorrect reception

Fig. 4 shows the received DRM field strength at fixed locations versus the distance to the transmitter, in order to check the coverage range. The values of the locations where satisfactory reception was obtained are green coloured while the remaining ones are coloured red. Thus, a Simulcast DRM service coverage radius of 9.5 km from the transmitter can be observed for static reception, when transmitting only 0.25 kW.

In addition, the predicted field strength level according to ITU-R Recommendation P.368-9 [6] for ground-wave propagation is depicted using a discontinuous blue line. The parameter values for the prediction were 16 mS/m for conductivity (average value in centre of Spain) and 6 for the relative permittivity.

When locations measured in Pozuelo de Alarcón are compared with Madrid locations, unequal attenuation values caused by the different urban densities of the two environments are confirmed. The difference between the measured field-strength levels at Madrid locations and the ITU-R predictions can be up to 40 dB against the 15 dB of Pozuelo. The main causes of reception failure were low field-strength values due to the urban reception environment and the distance to the transmitter. Man-made noise was an additional factor.

With regard to mobile reception, overall results are shown in *Table 4*. The percentages of correct audio included have been calculated from the total number of road stretches and not by time interval. This clarification is important, since it can be seen that the majority of the routes have higher percentages than 90%. However, if a value of 98% for correct audio subframes is set as the threshold, the percentage of stretches with satisfactory reception decreases significantly. Therefore, the power of the DRM Simulcast component is not enough to provide full mobile coverage in Madrid.

Table 4
Summary of routes of DRM in urban environment

Total Stretches	34
Total Kilometres	88
% Stretches with AudioQ > 90%	70.6
% Stretches with AudioQ > 98%	35.3

The East Radial runs from the transmitter to the outskirts of Madrid across the city centre. The measured AudioQ of its sub-stretches is depicted in *Fig. 5*. Green points represent 30-meter sub-stretches with 100% correct audio frames, while red points are 30-meter sub-stretches with lower AudioQ values. In this way, any sub-stretch that did not feature perfect reception is highlighted in red. This depicting procedure is very restrictive since correct reception was subjectively perceived along many red 30-meter sub-stretches.

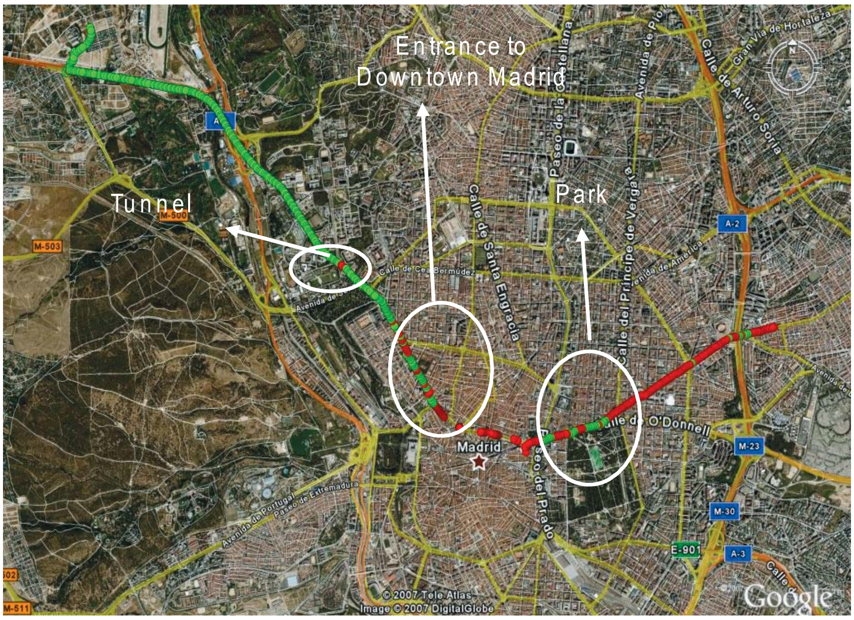


Figure 5
DRM signal AudioQ every 30 meters along the East Radial route

The beginning of the route goes from Pozuelo de Alarcón to the western side of Madrid centre. It is a suburban environment where reception was almost perfect: high field-strengths and SNR (MER) values were measured. However, as soon as the dense urban surroundings of Madrid were approached, the AudioQ of the sub-stretches did not fulfil the established criterion for correct reception. Only along wide streets with low buildings in the centre of Madrid (for instance, around the park) did the AudioQ return correct values. Mobile reception of the Simulcast DRM service in urban environments is limited by the lack of signal level.

AM-DRM mutual influence

This section analyzes the compatibility of the AM and DRM Simulcast components, taking into account different back-off ratios and receivers. *Table 5* shows the percentages of locations providing correct reception of the AM service and the four Simulcast configurations that were tested. Each receiver was individually analyzed.

Table 5
Quality of the AM Simulcast service with configurations MCS 10001, 10002, 10003, 10004

Mean Distance to Tx = 8.8 km	MCS 10001 (D = 16 dB)	MCS 10002 (D = 11.8 dB)	MCS 10003 (D = 7.2 dB)	MCS 10004 (D = 1. 5 dB)
Receiver	% Reception OK	% Reception OK	% Reception OK	% Reception OK
A (High-end)	100	100	60	100
B (High-end Mid-range)	100	50	0	0
C (Mid-range)	100	50	20	20
D (Mid-range)	100	50	0	0
E (Mid-range)	0	0	0	0
F (Mid-range)	20	0	0	0

The high-end receiver provided good results for each configuration, although the values corresponding to the MCS 10003 and 10004 configurations are close to the threshold. Similar results were obtained with receiver B.

The remaining mid-range receivers presented different behaviours. Receiver D provided little worse behaviour than Receiver B, which gave very good results with the MCS 10001 configuration, and results that were close to the threshold with the MCS 10002 configuration. Receiver C presented a similar behaviour to the two previous ones. In the case of the MCS 10003 and 10004 configurations, no satisfactory results were obtained with the mid-range receivers. Finally, the results for Receivers E and F were poor for all the MCS configurations.

Some of the worst results for the MCS 10001 configuration AM service were recorded at one particular location with very high and constant back-off ratios and AM signal values. This result proves that incorrect reception of the AM Simulcast component was not due to the lack of signal power or the environment. Therefore, the conclusion is that this mid-range digital receiver had problems with the AM service because of the DRM Simulcast component.

Monocast DRM tests

The second stage of the field trials was carried out while broadcasting only the DRM signal, without any analogue signal. This mode will be referred as Monocast in this article. In this case the DRM signal power was increased to 10 kW and the central frequency was 1260 kHz.

The measurements carried out in this campaign were planned in the city centre of Madrid, in order to determine the behaviour of the DRM system in a dense urban environment. Four areas were chosen for the measurements:

1) the *East of the city* with wider streets than the city centre;
2) the *centre of downtown Madrid* which comprises narrow streets with quite high buildings;
3) the *South of Madrid*, a representative environment for Spanish cities in general, characterized by irregular building heights (8-10 floors) and medium width streets;
4) Finally, the fourth environment was a *village located in the South-West of the city*, showing some common characteristics of types 1 and 3.

Table 6 summarizes the measured locations and mobile routes.

Table 6
Measurement summary

Zone	Type	Quantity
1. Salamanca	Points	13
	Routes	2
2. Gran Vía	Points	16
	Routes	17
3. Carabanchel	Points	15
	Routes	16
4. Vallecas	Points	8
	Routes	4

Global results for the static reception tests are shown in *Table 7*. Only one measured static location provided incorrect reception with an AudioQ value near 41.6%. The remaining 51 locations fulfilled the 98% quasi-error-free reception criterion.

Table 7
DRM reception quality – global results

Measured mode	9kHz/A/64/16/0.6/S
Total number of measured locations	52
Number of locations with audio quality better than 98%	51

In spite of the high man-made radioelectric noise levels typical in a big city, signal-to-noise ratio values greater than 20 dB were obtained, except for the problematic reception point mentioned above.

Table 8
Routes summary – mobile reception

Routes (Total)	39
Kilometres (Total)	133
% Routes with AudioQ > 98% (total route)	79.49

The results in *Table 8* refer to mobile reception. The majority of the routes had higher AudioQ values than 98%. The results obtained agree with the expected ones according to propagation in the medium-wave band. The results were influenced by critical factors that affect services in this waveband, such as the width of the streets, the height of the buildings (with regard to the width of the

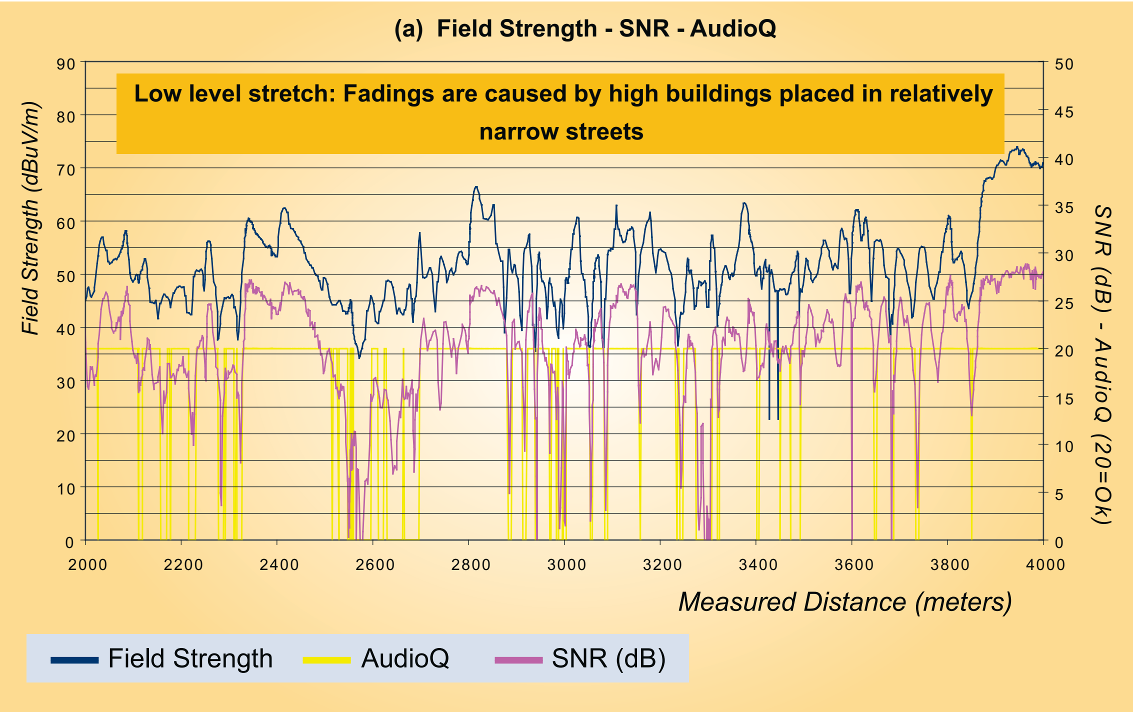


Figure 6
Salamanca district, east of the M-30: Field strength – SNR – AudioQ vs distance

corresponding street), man-made noise sources and urban elements that cause attenuation of the received signal (bridges, footbridges, tunnels….). These factors become critical as the signal level decreases with increasing distance from the transmitter. It occurs in the Salamanca zone shown in *Fig. 6*.

Fig. 6a (on the previous page) shows the electric field-strength level, SNR (MER) and AudioQ versus distance while *Fig. 6b* shows the reception quality calculated for 30-meter sub-stretches.

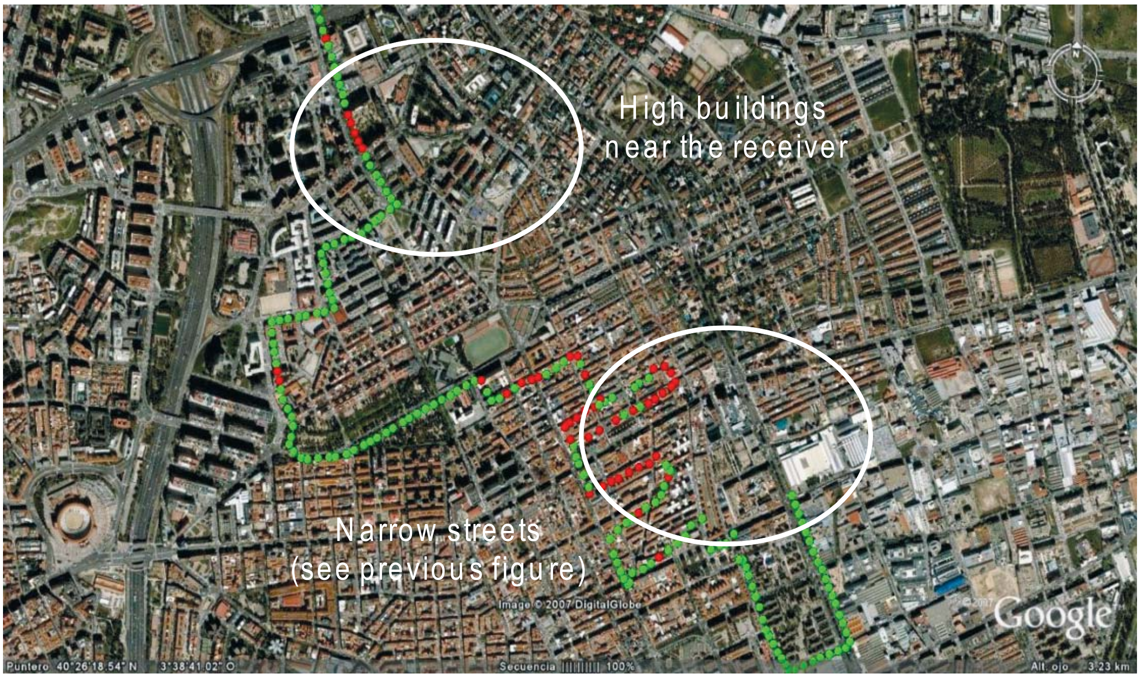


Figure 6b
Salamanca district, east of the M-30: reception quality over 30-meter sub-stretches

An increase in the broadcast power would slightly improve the signal reception in any of the analyzed cases. However, as shown in *Fig. 7*, most of the signal fading occurrences are too great to be rectified by a reasonable increase in the transmission power. The graphic shows that only one point returned an AudioQ value below 98%. This point is located slightly lower than the 17.5 dB SNR level. However, taking into account detailed results observed in each zone, a more conservative value of 18 dB is recommended for the SNR threshold in the case of static reception.

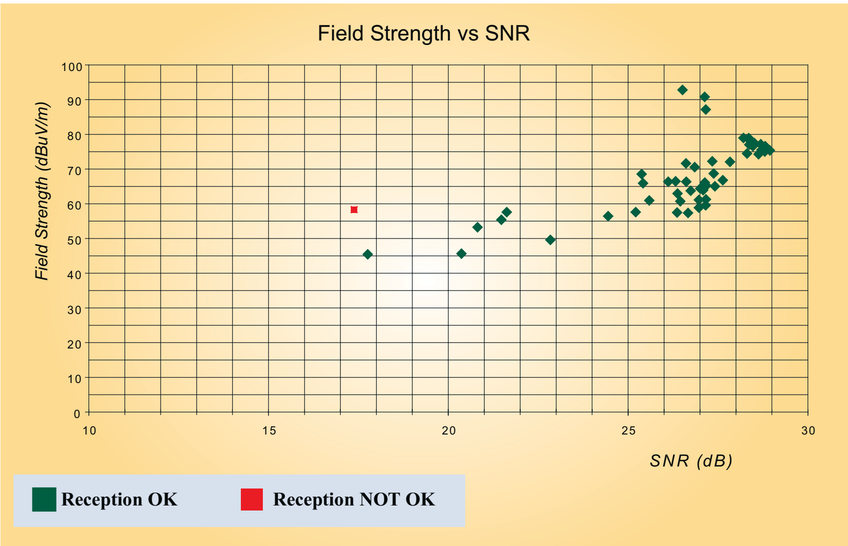


Figure 7
Electric field vs. SNR at fixed points for DRM monocast (10 kW)

In the case of mobile reception, the threshold value for DRM network planning in urban areas, broadcasting in the 9kHz / A / 64QAM / 16QAM / 0.6 / S mode, should be **20 dB**.

Both of the SNR thresholds obtained experimentally during these tests differ from the value of 15.1 dB proposed by ITU-R Recommendation BS.1615 [4] (which is based on laboratory tests and simulations).

These SNR values do not have a clear equivalence to the received electric field strength, because the electric noise levels measured during these tests were very variable. As for man-made noise, higher levels than the ITU reference [7] for urban environments were generally measured. Noise levels of around 30 dBμV/m were measured, showing a variability of up to 8 dB in some areas.

Another important factor is the level of attenuation along a route, which depends on different local urban factors. The median variation between the average value and the value which was exceeded at 99% of locations was 22 dB (standard deviation = 7 dB).

Finally, a study to determine the field strength spatial distribution at a lower frequency in urban environments was performed. For this purpose, a 50 kW radiated AM signal was broadcast on 810 kHz from the same transmitting site used for the DRM experimental network. The same routes were measured for both signals (810 kHz and 1260 kHz) and a received field-strength comparison was made after normalizing the levels measured at both frequencies.

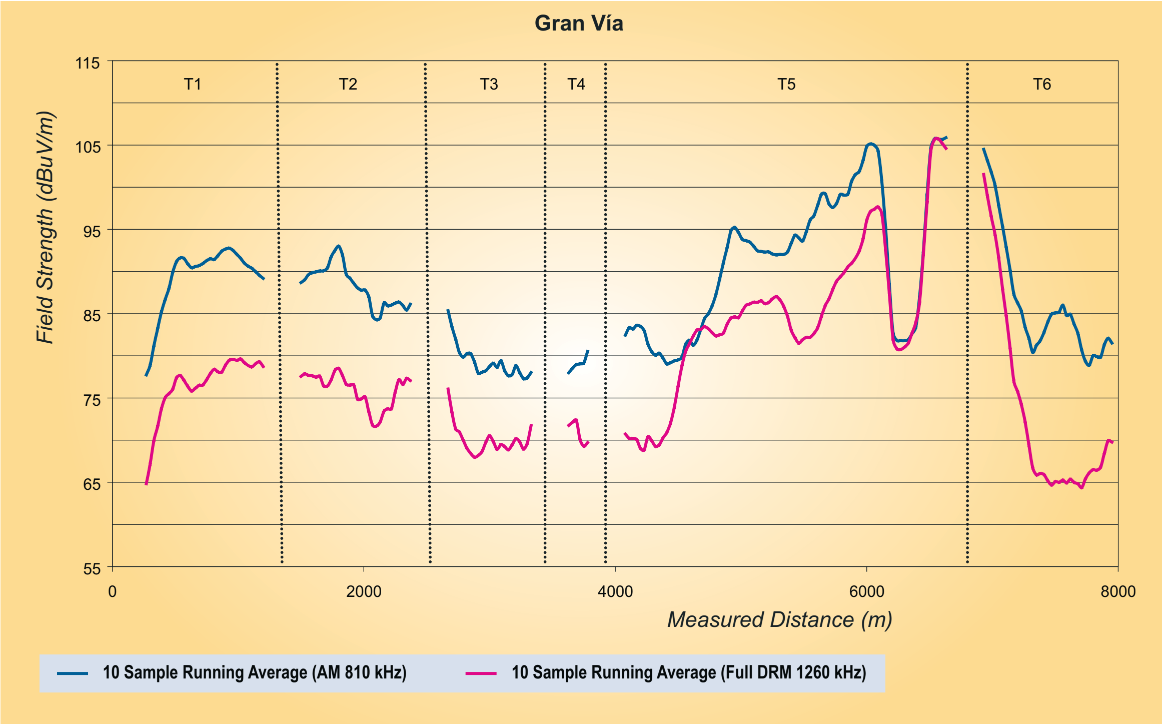


Figure 8
Gran Vía zone. Received field in 810 kHz and 1260 kHz

By means of this procedure, almost 40 km of routes were assessed in the different zones covered by these tests. *Fig. 8* shows the received field strength for six sections analyzed in the Gran Vía zone. In order to make the comparison easier, a ten-samples running mean was applied, reducing the fluctuations of both signals.

It can be observed that both of the signals vary in a similar way with a 10 dB level difference. The difference in the received field-strength level for both of the studied frequencies in the whole city of Madrid is determined by averaging the global values obtained for all areas as shown in *Table 9*.

According to these results, it can be concluded that the received field strength levels on 810 kHz in the city of Madrid are generally 10 dB higher than the ones received on 1260 kHz. However, this is not a constant difference, as it can fluctuate depending on the urban elements of the environment so that the difference between the two signals increases when the height and density of buildings is higher, whereas the opposite effect takes place in less dense environments. Taking into account that the wavelength corresponding to 1260 kHz is shorter than the wavelength for 810 kHz, the former frequency is more affected by surrounding obstacles. It is worth pointing out that the distance between the transmitter and reception location is not a decisive factor in that difference.

Table 9
FS differences between 810 kHz and 1260 kHz statistics

Zone	Mean value (dB)	Standard Deviation (dB)
Salamanca and West M-30	12.08	5.42
Gran Vía	10.00	5.95
Carabanchel	8.96	4.48
Vallecas	8.24	4.49
MADRID CITY	9.82	5.08

DRM Monocast indoor tests

During the last 5 years, several studies have been carried out based upon field measurements in the medium-wave band. Different modes and configurations, correct reception thresholds, interference and environments have been assessed but this is the first time that an extensive specific field trial has been carried out for portable indoor reception analysis in the MW band. This environment was assumed to be more hostile and critical, mainly due to an increase in man-made noise levels.

The measurements for this stage of the field trials were planned for 6 buildings in the city of Madrid and one in an industrial area in a suburb to the south. With the aim of determining the performance of the DRM signal, different types of buildings were chosen, in different environments and serving different purposes. A total of 113 indoor locations in seven buildings were measured as summarized in *Table 10*.

Table 10
Measurements summary

Identifier	Type	Streets	Height (floors)	Environment	Distance to Tx (km)	Number of meas. points
E1 Building	Private	Narrow	3	Urban Non Dense	13.8	15
E2 Building	Private	Narrow	3	Urban Non Dense	10.2	9
E3 Building	Private	Wide	7	Urban Dense	7.9	18
E4 Building	Private	Wide	10	Urban Dense	13.2	8
E5 Building	Commercial	Wide	10	Urban Dense	8.9	44
E6 Building	Commercial	Wide	2	Industrial	16.9	11
E7 Building	Commercial	Wide	6	Urban Dense	9.9	13

The overall results for indoor reception are displayed in *Table 11*. These results have been divided according to the environment where the building is placed and the type of building. DRM reception in the medium-wave band depended on the height of the buildings and its adjoining environment. Noise values of 51, 64, 64 and 62 dBμV/m were obtained in four different dense urban environments. For planning purposes, this parameter has to be considered as essential owing to the field level requirement. This value should be as high as 80 dBμV/m for satisfactory reception in dense urban environments, while it can be reduced by more than 10 dB in non-dense urban environments.

Table 11
Summary of evaluated buildings (median values)

	PRIVATE BUILDINGS				COMMERCIAL BUILDINGS		
	E1	E2	E3	E4	E5	E6	E7
Corrects (%)	80	100	56	0	25	64	62
E (dBμV/m)	78.95	93.79	74.31	68.08	79.51	70.8	83.99
SNR (MER) (dB)	25.04	27.17	20.44	5.08	15.25	21.19	20.93
N (dBμV/m)	43.81	53.32	51.54	64.61	64.17	52.72	62.13
Dist. Tx (km)	13.8	10.2	7.9	13.2	8.9	16.3	9.9
Environment	Urban Non Dense	Urban Non Dense	Urban Dense	Urban Dense	Urban Dense	Industrial	Urban Dense

As for the noise behaviour, for instance, *Fig. 9* shows the effect of an elevator on the noise level and therefore over the received SNR value which, in this case, is near the correct reception threshold value.

As a conclusion, 3 out of 7 buildings offered acceptable indoor reception. Good reception could be obtained in the higher part of the building, that is, on the second or third floor of a three- or four-storey building. The best reception reliability was found in non-dense urban environments. Inside the remaining measured private homes, E3 and E4 only showed good reception near the windows. That is because they are very tall buildings located in dense urban environments.

At the headquarters of *Cadena SER* (E5), reception was not good in general terms, and depended on the floor where the evaluation point was placed. SNR values showed variations from 0 to 17 dB. As in the previous case, and because of being placed in a dense urban environment, the reception close to the windows was better. In this special case, the analogue services transmitted on 810 kHz, were also not correctly received**.** The main reason for bad reception in this building is man-made noise. Noise levels can fluctuate up to 40 dB, with a median value of 19 dB.

What is also remarkable is the variation of the received electric field strength inside the buildings: they can vary by up to 30 dB inside the same building. The median value of the measured variation was 16 dB. Moreover, this value increases with the height of the reception point.

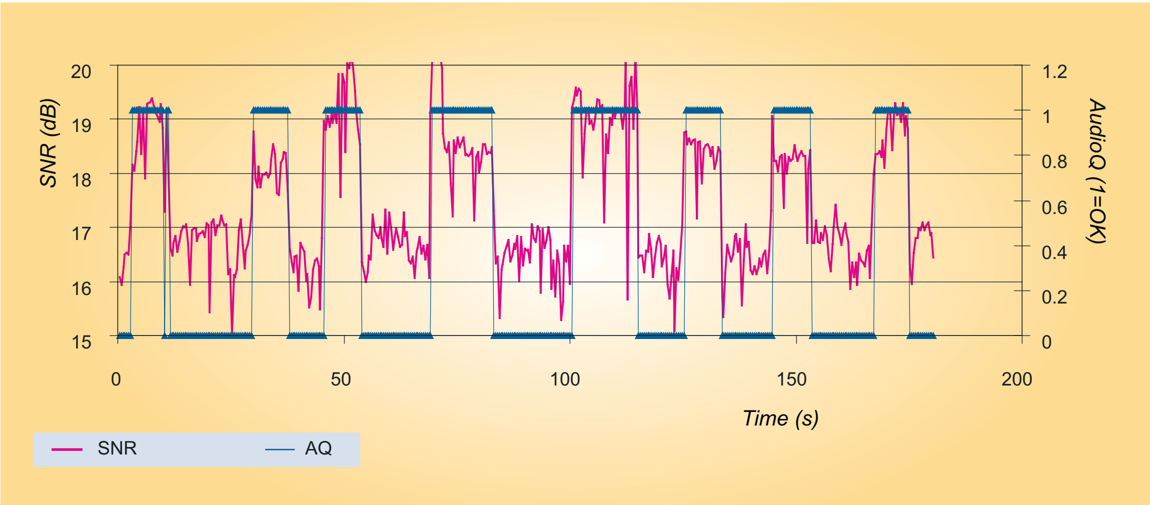


Figure 9
Threshold levels in E5 building

The required signal-to-noise ratio for a correct reception of the DRM signal using the mode mentioned above was 17.5 dB. Using 10 kW transmitted power, around 48% of the measured locations were covered. As a conclusion, this transmission power is not enough to provide reasonable indoor reception in a big city like Madrid. However, according to the results, with an increase of 3dB in the SNR value; i.e. increasing the transmitted power by 3 dB (equivalent to a 20 kW transmitted power), a coverage of 82% of locations can be achieved. *Fig. 10* shows the evolution of the predicted coverage when a transmitted power increase is considered. The 0 dB bars correspond to the actual coverage measured in these trials with 10 kW, so +3 dB would be 20 kW and +6 dB would be 40 kW. Each row represents each building tested in these trials (E1 to E7)

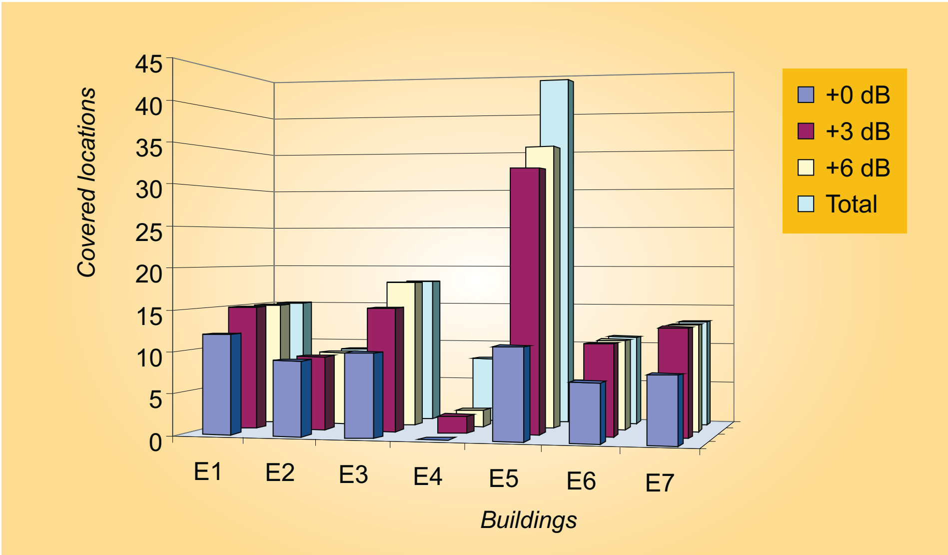


Figure 10
Coverage prediction

The first row (foreground from left to right) represents the number of indoor locations with good coverage (AQ>98%) inside each building. The second row shows those which would be covered by increasing the transmission power 3 dB, that is, to 20 kW. The third row shows the hypothetical case of a 6 dB increase (to 40 kW) and, in the last row, the total number of points measured in each building is shown.

The actual test transmission, featuring 10 kW transmitted power, only covered the E2 building in its entirety. However, in a hypothetical case of a 3 dB increase in transmission power (to 20 kW), the coverage would considerably increase to 100% in the E1, E2, E6 and E7 buildings. Good coverage within the E3 building (83%) and E5 building (73%) would also be obtained.

In the case of a 6 dB increase in transmission power, the coverage will improve only a little in comparison with the previous prediction, with the E3 building being totally covered. The E4 building and a low percentage of the E5 building would not be covered, due to local propagation conditions. Even if the transmission power were increased, it would not enable correct reception of any type of analogue or digital signal in the medium-wave band.

Conclusions

A DRM experimental network was set up in Madrid during 2007 to evaluate the AM-DRM Simulcast configuration for outdoor environments and the DRM Monocast configuration (in both urban outdoor and indoor reception environments).

The coverage area when broadcasting in the ***Simulcast configuration*** resulted in around 7 km with 16 dB power ratio and 0.25 kW transmitted power for the digital signal. This back-off ratio value is mandatory for the worst performing receivers. On the other hand, the analogue signal should be given a minimum of 12 dB margin over DRM in order to get acceptable reception with most mid-range receivers.

In the case of ***DRM Monocast reception in urban outdoor areas***, the transmitted power of 10 kW provided an excellent static coverage. Nevertheless, a transmitted power of 20 kW would be needed to achieve fair mobile coverage. Also, low frequencies within the medium-wave band ensure lower attenuation values in these environments, thus extending coverage significantly.

Finally, the results for ***DRM Monocast indoor reception*** depend highly on the type of building. Reception is generally better in the upper part of such buildings and Non-Dense urban environments. Since high noise levels were found at many locations, the transmitted power should be increased to 25-35 kW to cover Madrid properly.

Acknowledgements

This work has been a collaborative effort, following an initiative from Cadena SER, with the network operator AXIÓN, the transmitting equipment manufacturer VIMESA and the University of the Basque Country (UPV/EHU). The work was partially financed by the Spanish Ministry of Industry, Tourism and Trade under the PROFIT programme as “DETEC DRM”.

David Guerra received an M.Sc. in Telecommunications Engineering from the University of the Basque Country, Spain, in 2000. He received a Ph.D. in September 2006 following a study of the planning parameter values for Digital Radio Mondiale services propagated over the ground in the medium-wave band. In 2002, he joined the TSR (Radiocommunications and Signal Processing) research group at the Department of Electronics and Telecommunications of the University of the Basque Country, where he is currently an Assistant Professor. His research interests focus on radio propagation and network planning for digital broadcasting services.

Unai Gil received an M.Sc. in Telecommunication Engineering in 2004 from the Bilbao Engineering College, University of the Basque Country. He has been involved in several research projects during the last four years in the area of digital terrestrial TV and digital radio broadcasting with the University of the Basque Country. His current research interests include several aspects of Digital Radio Mondiale signal processing applied to a Simulcast configuration. Nowadays he is working towards a Ph.D. thesis on this subject.

Luis del Amo received a degree in Telecommunications Engineering from the University of Alcalá de Henares, having specialised in Electronic Equipment. He has a High Degree in Telecommunications Engineering from the Polytechnic University of Madrid. In 1979 he joined the Radio and Television Engineers Corps by Public competition. He was appointed Deputy Technical Director of Cadena SER in 1989 and has been taking part in all the projects on new technologies at Cadena SER.

Juan Masdeu Garcés graduated from the Polytechnic University of Madrid with a degree in Telecommunications Engineering, having specialised in the field of "Image and Sound". In 1982 he started working at Radio 80 in the Radio Frequency section before moving one year later to Cadena SER as Head of the Technical Department in Catalonia. Later, he moved to the headquarters of Cadena SER, becoming responsible for Radio Frequency matters. He is also in charge of planning and development of local stations across the Cadena SER network.

The authors would like to specially thank TRANSRADIO for providing the transmission equipment and the technical support that was necessary during the field trials.

Bibliography

[1] Digital Radio Mondiale: **DRM Simulcast test report to ITU-R V1.0**
ITU-R WP 6E 199E, March 2002.

[2] Gil U. et al.: **DRM 20 kHz Simulcast Field Trials in the Medium Wave Band in Mexico D.F.**
IEEE Transactions on Broadcasting, Vol. 54, March 2008.

[3] TS 201 980 V2.1.1: **DRM - DIGITAL RADIO MONDIALE**
ETSI Standard, 2004.

[4] ITU-R Recommendation BS.1615: **Planning parameters for digital sound broadcasting at frequencies below 30 MHz**
International Telecommunication Union - Radiocommunication Sector, 2003.

[5] ITU-R Recommendation BS.1284-1: **Methods for the Subjective Assessment of Sound Quality: General Requirements**
International Telecommunication Union - Radiocommunication Sector, 2003.

[6] ITU-R Recommendation P.368-9: **Ground wave propagation curves for frequencies between 10 kHz and 30 MHz**
International Telecommunication Union - Radiocommunication Sector, 2006.

[7] ITU-R Recommendation P.372-8: **Radio noise**
International Telecommunication Union - Radiocommunication Sector, 2003.

Digital Radio

— a receiver manufacturer's viewpoint

Colin Crawford
PURE Digital

Tony King-Smith
Imagination Technologies

While DAB take-up has been very successful in a few markets, it has failed in other markets due partly to a lack of compelling digital-only content.

In this article, PURE Digital – the most successful DAB receiver manufacturer over the past five or six years – provides an upbeat analysis of what the future holds for DAB-based digital radio. The article concludes with a section from parent-company, Imagination Technologies, on the latest silicon developments for digital radio.

The first non-experimental DAB digital radio transmissions began simultaneously in the UK, Denmark and Sweden in Sept 1995, instigated by the BBC, Danmarks Radio (DR) and Sveriges Radio. The expectation back in those days, that radio listeners would be impelled to buy DAB radios largely on the basis of improved audio quality proved to be misplaced, resulting in a lack of demand by consumers and a resulting lack of affordable receivers from radio manufacturers. In fact, for the next seven years, the only DAB receivers were Hi-Fi tuners (with the exception of two extremely expensive "portable radios" which were effectively Hi-Fi tuners with attached speakers), appealing only to the relatively affluent audiophile market.

It was only when the BBC and commercial radio in the UK realized that content was the key driver of digital radio listening, and separately started transmitting unique DAB content, that a viable market could be envisaged, and development started on new DAB chipsets which would allow affordable DAB radios in various mainstream form-factors (kitchen, bedside etc.). The first tangible result of these developments was the development of the Chorus processor by Imagination Technologies. It was then licensed to Frontier Silicon and brought to market as the Chorus FS1010 chip. The Venice module produced using this chip was used in the EVOKE-1 – the world's first sub-£100 digital radio – manufactured by PURE Digital, a completely new digital radio brand.

The EVOKE-1 started shipping in August 2002 and, by the end of that year, it had single-handedly trebled the installed base of DAB radios in the UK. More importantly, it applied pressure on the incumbent radio suppliers in the UK, who saw over the next year a significant loss of market share, and forced them to bring forward DAB product plans which otherwise had very little chance of being completed.

By the end of 2003, a number of other new radio brands had entered the market – having seen that digital radio provided a means to make inroads into what had previously been a market dominated by a few very traditional brands. Products from these new manufacturers – eventually along with new products from the traditional manufacturers – meant that the uptake of DAB digital radio, when viewed with August 2002 (rather than September 1995) as its starting point, has subsequently moved along a very standard successful technology introduction curve, in line with CD, DVD and others.

Looking at the UK audio market as a whole (combining radio with the burgeoning MP3 market) at the peak Christmas selling season, it would come as no surprise to know that the five top-selling products are all Apple iPods. What might surprise however is that after just one other product (a very inexpensive MP3 player), the next best-selling audio product is a portable DAB radio – the PURE ONE. And the next DAB radio appears after only another two iPods and a Samsung MP3 player. In fact, if Apple is excluded from the equation (effectively treating Apple as a market in its own right, which it effectively is), the DAB and MP3 markets in the UK have very similar values – £162M for DAB and £193M for MP3 for the year ending July 2007.

DAB, where it is successful, is **very** successful.

DAB geographical markets – a mixed bag

Successful DAB markets, and the reasons for growth

It is well known that the stand-out market for DAB digital radio is the UK, with around 7 million receivers now being used in approximately 27% of UK households. What is less well known is that a small number of other countries have seen similar success when viewed in proportion to the population.

Denmark in particular has a thriving DAB market, with a household penetration almost identical to that of the UK. Norway follows close behind Denmark, with Belgium and Switzerland not far behind that.

> ***“ Denmark in particular has a thriving DAB market ”***

So what is it about these countries which has created successful markets where other countries, notably France, Germany, Spain, Italy and Portugal have so far failed to generate any tangible success? What are the common elements which have combined to create successful market conditions?

Progressive regulation

Successful DAB countries tend to have regulators which have recognized the importance of DAB digital radio to the future of broadcasting within their jurisdiction, and have created regulatory regimes which have encouraged interest and investment from broadcasters. This has variously included automatic extensions of FM licences plus requirements on broadcasters to actively promote digital radio.

Enthusiastic public service broadcasters

The PSB in each of the successful countries has been prepared to lead the way on DAB transmission, as well as take strong measures towards promoting DAB on the platforms available to them.

They have also in the main produced unique digital content, seeing that as the key driver for digital radio up-take.

Commercial radio involvement

Each of the successful countries has a high PSB share of listening, but commercial radio is an essential part of the digital radio mix. Where commercial radio has been encouraged or been visionary enough to get involved, this has greatly accelerated demand, partly because the very direct digital radio promotion by commercial broadcasters tends to complement the more generic and broader promotion from PSBs.

> ***“ Each of the successful countries has a high PSB share of listening, but commercial radio is an essential part of the digital radio mix ”***

Unique digital content

Of the four key selling points of digital radio (content, ease of use, extra features and improved sound quality), the one which consistently tops surveys of why we should purchase a DAB radio is **content**. The BBC has produced six digital-only stations, one of which (BBC7) is consistently mentioned as a key reason to buy a digital radio. Danish Radio has produced a complete bouquet of unique (essentially repackaged) services around the concept that people want to listen to their particular type of radio station exactly when they want to, not when the broadcaster decides to broadcast it. This model has been repeated in Norway and is also being adopted in Ireland. Even commercial broadcasters have produced unique content, generally focusing on strong niche markets. Some of these stations (e.g. Planet Rock in the UK) have taken significant market share.

Broadcaster promotion

Consumer demand is driven by promotion and in the successful countries where the broadcasters see that their future is assured – not threatened – by digital radio, this promotion is coordinated and takes place. Promotion drives awareness and demand and that in turn drives the retailers to buy, stock, display and again promote products. In this way a vicious circle [1] is changed into a virtuous circle of demand and investment.

Stagnant DAB markets, and the reasons for slow growth

> ***“ There are over 300 DAB receivers available, many from reputable brands ”***

As little as three years ago it might have been reasonable to blame the failure of some markets on a lack of available and affordable DAB receivers, but that is clearly no longer the case. There are over 300 DAB receivers available, many from reputable brands, covering a very wide range of price points starting from as low as £25 (32 Euros), and receiver manufacturers are keen to make their products available in as many markets as possible. So why aren't they readily for sale in the countries mentioned above?

1. No unique selling points (especially content) = nothing to promote = no promotion = no demand = no product in the shops = no reason to invest in DAB = no unique selling points…

Some of the reasons are of course the opposite of those above:

- In Spain for example, a seemingly reasonable attempt by the regulator to share out the available DAB spectrum evenly among the national broadcasters meant that the larger broadcasters are under-represented in the digital space, and are therefore reticent to promote DAB. The smaller broadcasters, although keen to promote DAB for the very same reason, have little or no analogue platform on which to do so.
- German public broadcasters have generally been less than enthusiastic about digital radio and, with a few exceptions, have not invested in producing unique digital content.
- Digital radio legislation and regulation in France was not given a necessarily high priority, and so took a very long time to come to fruition.
- In Italy, the job of promoting and lobbying for DAB digital radio sits not with a single body, but with a number of commercial radio groups, each with their own agendas and issues. Rather than coming together in the interests of creating a successful market which is then worth fighting over, the groups have tended to fight among themselves, leaving the digital radio market open for competitive technologies. Fortunately, recent news seems to indicate that a new era of cooperation has started in Italy, and that bodes much better for the future.

Some other reasons also exist however.

Failure to understand the radio receiver market

Broadcasters are at a point removed from the business of manufacturing and selling receivers and have at times put measures in place which although seemingly helpful, have in fact been negative. For example large-scale give-aways and broadcaster-subsidised radios has had the effect of stifling retailer demand. In fact, although broadcasters and manufacturers have the same ultimate goal (millions of receivers in the market), their preferred means and timescale for getting there are fundamentally different. Broadcasters would like to see £10 receivers in the market immediately. The major manufacturers on the other hand, have invested millions developing chipsets and radios and they need to see financial returns on that investment. For the manufacturer, a key part of this is to avoid a price crash which would result in minimal (or potentially negative) profit margins and ultimately in no incentive for the market to exist. This is almost a catch 22 situation, but not quite. The result is that it just takes longer than some might like for the market penetration to build.

Low penetration of in-car DAB receivers

It is undoubtedly true that the lack of in-car receivers has slowed down the acceptance of DAB. A large proportion of European radio listening is done in the car, and the long development cycles of car manufacturers, hindered by the heterogeneous nature of the market (see below), have been a major dampener. Adapters to retro-fit DAB into existing cars and after-fit radios have so far failed to take off, but the new PURE Highway – an affordable in-car DAB adapter which can be installed like a sat nav in a couple of minutes *(see the photos below)* – looks to be making a difference. Using FM

retransmission to get DAB radio onto the existing car stereo, while also supporting iPods and the ability to skip back to something you missed, has definitely created a jump in this sector and might prove the catalyst it requires while the car manufacturers get their plans in place.

Heterogeneous and confusing geographical market

The in-car issue above is compounded by the very different DAB experiences throughout continental Europe. Whereas an FM listener can drive from country to country and know that there will be content available, the same is not true of DAB digital radio.

Non-global market

All technologies require investment and in most cases such investment can only be justified within the context of a global market. While DAB digital radio remains successful in only a few markets it will not get the attention of the mainstream consumer electronics manufacturers – with a subsequent lack of investment, marketing and slower growth.

DAB – what sells and to whom

> ***“ ... what types of devices are selling, and who are the people that are buying them? ”***

Despite the doom and gloom of the previous section, investment is taking place in this sector and the market is growing, albeit slowly in most countries for now. But what types of devices are selling, and who are the people that are buying them? Most of the relevant statistics come from the UK.

DAB consumers have principally been male biased *(see Fig. 1, top left)*, relatively mature *(Fig. 1, top right)* and coming from the higher end of the social scale *(Fig. 1, centre left)*. Some reasons for this customer profile are as follows:

1) DAB radios are (and will remain for some time) more expensive than FM radios, and so purchasers of DAB radios are likely to be the people who have a real passion for radio. That tends to be the older generation who were brought up with radio (as opposed to TV).
2) DAB products have been focused at that demographic for the simple reason that those are the products which sell best. This could be thought of as another chicken and egg situation, but it is also true that most DAB products aimed at a younger audience have fared less well in the market.
3) There is a school of thought within marketing which says that sub-£100 is the "sweet spot" for a consumer electronics product for a man – the price at which it becomes a reasonably impulsive buy. That price is thought to be sub-£50 for women.

This reasoning is borne out quite strongly by PURE’s experience of selling its classic wooden radio products (strongly AB social scale, strongly male, strongly mature), versus the experience of selling the more mainstream-styled and affordable (£49.99) PURE ONE. ONE still sells to the same demographic, but broadens it considerably to include many more women, C1 and even C2 social scale, and a slightly younger audience (perhaps helped by the presence of a pink version).

ONE has also sold over a quarter of a million units in less than 18 months, something which took over three years to happen with the current best-selling DAB radio – the EVOKE-1S (and its predecessors, the EVOKE-1 and the EVOKE-1XT). Although the DAB market was hailed as going mainstream when the millionth receiver was announced back in December 2004, it is actually the arrival of sub-£50 radios (effectively sub-70 Euro radios) and the 5 million sales point achieved in May 2007 which heralded a real mainstream market, and allows us to look impassionedly at the types of receivers which actually sell in volume and the niche markets arising within the overall DAB receiver market.

No impartial figures exist to show the split between “basic” DAB radios and “advanced” DAB radios, and so we have had to look at the sales of PURE products in isolation. However, given that PURE has produced the most commercially successful advanced radios, the conclusions below are likely

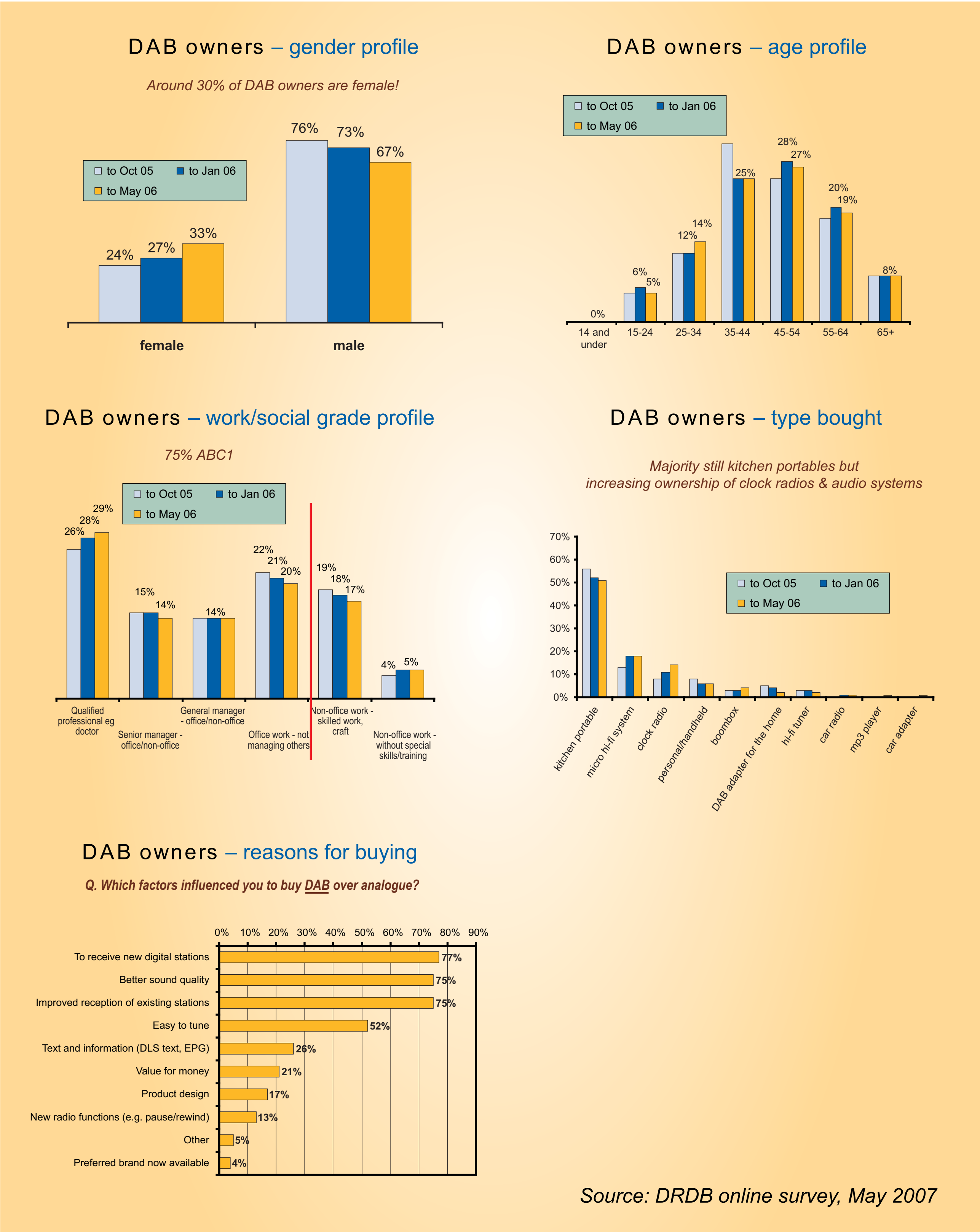
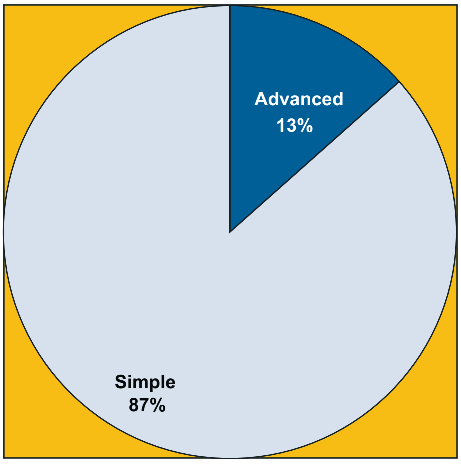


Figure 1
Some reults from a DRDB user survey, May 2007

to be only strengthened if the whole market was taken into account.



In this analysis we are defining "simple" radios as those without advanced features such as an Electronic Programme Guide, pause & rewind and record to an SD card.

Looking back over the last year, "simple" radios as defined above made up almost 87% of the sales of all PURE DAB products sold, leaving 13% of sales accounting for "advanced" receivers *(see pie chart)*. Analysing this by radio category, it appears that people want simple bedside radios, some want advanced features in a kitchen radio, but many want extra features in a micro system, perhaps because it is a form factor which they want to think of as "future-proof".

New DAB categories

DAB has undoubtedly been successful in the portable radio category, often called kitchen radios, to the point where the vast majority of the value of that market is DAB radios, even persuading some retailers to stop selling analogue radios. The same is true to a slightly lesser degree in the personal (handheld) category. But the real growth categories for 2007 have been the mini/micro system market, and particularly the clock radio (bedside) market. These are burgeoning on the back of the launch of excellent products from mainstream manufacturers. But looking out to 2008, there will be a few key categories which should see significant growth:

Combined DAB & iPod docks

The iPod docking market is still relatively small, but is rapidly expanding. Chronos iDock became the best-selling docking radio in the UK after only 1 month on the market.

In-car DAB adapters with FM retransmission

The PURE Highway is the first of what is likely to be a new breed of in-car radios and adapters – providing the choice of DAB as well as some of the extra features of DAB for the first time in the car, and most importantly in a format which can be fitted by the least technical owner.

DAB & Wi-Fi radios

Wi-Fi internet radios have been available for some time but have been relatively unsuccessful to date. Radios are expected soon which will combine the broadcast strengths of DAB with the "connected" strengths of internet radios, allowing access to thousands of internet radio stations, on-demand radio stations and podcasts, but more critically also allowing consumer interaction with the radio station as well as a possible link between the local strengths of radio and the social networking sites that are today present on the web.

New features of DAB

Many new features have been introduced into DAB products since the EVOKE-1 was launched with its very simple functionality and interface.

Abbreviations

AAC	Advanced Audio Coding	**DVB-H**	DVB - Handheld
DAB	Digital Audio Broadcasting (Eureka-147) http://www.worlddab.org/	**EPG**	Electronic Programme Guide
DAB+	DAB using the AAC codec	**FM**	Frequency Modulation
DMB	Digital Multimedia Broadcasting http://www.t-dmb.org/	**PSB**	Public Service Broadcaster
DRDB	(UK) Digital Radio Development Bureau http://www.drdb.org/	**SD**	Secure Digital (memory card)
DVB	Digital Video Broadcasting http://www.dvb.org/	**T-DMB**	Terrestrial - DMB
		TPEG	Transport Protocol Experts Group http://www.tpeg.org/

Pause and rewind

Sometimes called "ReVu," sometimes "PausePlus", pause and rewind on DAB radios is relatively straightforward to implement, requiring some extra memory and processing, though the extra memory accesses can cause sensitivity issues if not designed carefully. Although there are quite a few products with this feature, it has not proven itself to be compelling to the consumer in the same way as it has with hard-drive-based digital TV systems.

Record to solid-state memory (normally SD card)

Perhaps the most successful advanced feature to date (especially when combined with pause, rewind and EPG), is the ability to set timed recordings of radio programmes – more so than immediate recordings of current content.

Electronic Programme Guide (EPG)

Consumers now commonly use an EPG on television to decide what to watch and what to record. DAB EPG, now in use widely in the UK and somewhat less outside the UK, is a much used feature, allowing consumers to find out what is happening on particular radio shows, and set timed recordings to listen to at their convenience. Radio, like many entertainment media, is becoming increasingly consumed "on-demand".

Enhanced DLS (Intellitext, Journaline and DLPlus)

DAB scrolling text (DLS) is a much-appreciated unique feature of DAB, but is of course transitory. Enhanced DLS systems address this issue in different ways, but with the same ultimate goal – to store the information in a user-accessible database to allow them to access the information at their leisure. Systems currently being used allow users to find news headlines, sports results, weather and traffic information as well as a host of other data. At this point, WorldDMB has standardized on Intellitext, but is still considering Journaline and DLPlus as more data-centric approaches. Intellitext has the advantage that it requires no access to the data channel within the DAB stream, and relies very simply on structure within the visible DLS. The other formats use data to deliver tagging information with the advantage that they can deliver more complex data, but at the cost of more complexity for the broadcaster.

DAB slideshow

Currently in the trial phase, DAB slideshow is increasingly being transmitted by broadcasters in the expectation that it will add a strong visual element to radio in the same way that most MP3 players can display images of the band being listened to. Unfortunately, the significantly increased cost of the colour display, and the associated extra cost of full-colour DAB receivers which could show such slideshows, means that this feature is likely to be implemented only on relatively expensive devices. However, such devices are likely to be attractive to the iPod generation, and so great hopes are held out for their arrival some time in 2009.

TPEG

Although part of the full DAB specification, TPEG has to date seen little support in receivers and subsequently little broadcast support. However, the BBC is now transmitting TPEG data and the 4Digital group (who won the licence for the second commercial national multiplex in the UK) also has plans to broadcast a TPEG service. This is of course of interest for in-car receivers and adapters, but is also increasingly being seen as of interest for static (in-home) receivers, where consumers could find out the traffic situation before deciding on a particular route to work.

New radio features associated with DAB

On top of these features which are driven by the DAB system, a number of other new features have been introduced on DAB radios, simply because they link well with a processor-based radio system. Some of these features are the SnoozeHandle (a touch-sensitive metal handle for bedside radios); the ability to go to sleep to one station and wake to another; USB upgradeability to allow new features to be added post-purchase; ChargePAK – a rechargeable battery pack – which has come about due to the relatively high power consumption of digital radios.

Advances in DAB silicon

Since the earliest days of DAB broadcasting, a number of key semiconductor manufacturers have been developing devices to enable the transition of DAB from low-volume trials to mass market adoption. However, it was the combination of silicon solutions delivering substantially above the minimum performance specified, together with a strong market push from a company truly focused on the success of DAB and backed up by all major broadcasters, that was needed to get the sales moving.

For any consumer product in an aggressive cost-sensitive market, high levels of integration are key to boosting the unit shipments needed to generate commercial momentum. Thus, semiconductor devices – in volume production – that are available to a wide range of manufacturers are essential for an emerging technology such as DAB.

> ***“ One of the most successful chips in the development of the DAB receiver market has been the FS1010 (Chorus 1) chip from Frontier Silicon ”***

One of the most successful chips in the development of the DAB receiver market has been the FS1010 (Chorus 1) chip from Frontier Silicon, incorporating patented intellectual property from Imagination Technologies (the parent company of PURE Digital). Current estimates show that more than 70% of all DAB receivers sold worldwide use this chip, mainly as part of one of Frontier's Venice family of DAB modules. Frontier Silicon continue to develop their silicon and module product families as the market for DAB radios continues to grow.

However, another market force has emerged that is also driving semiconductor devices for DAB. In South Korea, the T-DMB mobile multimedia service was launched towards the end of 2005. T-DMB uses exactly the same modulation, framing and signalling as DAB, but also specifies higher layers to enable the reliable transmission of video and audio content, as well as extensive data services. Since the higher layers are often implemented in software, this has created a significantly larger market for chips that are capable of demodulating DAB transmissions – the Korean market alone for T-DMB was expected to reach more than 5m users by the end of 2007.

Extending the market for chips that are capable of receiving DAB is essential to support the relentless drive for lower costs. Another factor now impacting this is the worldwide interest in mobile TV, using systems such as DVB-H, ISDB-T or MediaFLO. An increasing number of these chips now also support DAB demodulation, as the market moves towards multi-standard receivers capable of supporting all the new standards. Imagination Technologies is a leader in this area, with a growing number of partners licensing their unique software-defined demodulation solution. These multi-standard solutions will increasingly enable any mobile-TV-enabled handset to also receive DAB – a key factor in achieving the “holy grail” of DAB becoming a reality in mobile phones.

The extension of the DAB standard to DAB+ (ETSI TS 102 563) allows as much as 3 to 4 times the number of audio services to be carried within the same multiplex as the older DAB system. This has opened up new markets for DAB – initially Australia and Eastern Europe, but should also reinvigorate markets such as Switzerland and Germany. However, the AAC+ codec plus Reed-Solomon error correction used in DAB+ makes significantly greater demands on the receiver subsystem. Unless the chips put more of these functions in hardware, these standards become more expensive to deploy. Fortunately, a growing number of chipsets using technologies (such as the ENSIGMA UCC multi-standard technology from Imagination Technologies) are capable of managing these extensions. Putting these functions in hardware and utilising the latest semiconductor manufacturing processes is enabling silicon manufacturers to continuously reduce power consumption while increasing functionality, thus ensuring DAB and DAB+ receivers continue to progress down the cost-reduction curve.

Audio decoding is another area that is changing. In dedicated DAB receivers, the audio decoding is usually done within the receiver chip itself. However, as DAB starts to become a feature in mobile TV and other multimedia mobile platforms, the audio decoding is being handled by the media applications processor driving other parts of the platform. This not only reduces the cost of adding DAB to mobile platforms, but also encourages mobile handset manufacturers to add low-cost multi-

Despite an honours degree in Software Engineering from Imperial College London and an early career in software development, **Colin Crawford** moved to the dark side and has for the past 6 years been responsible for all aspects of marketing PURE’s digital radio products worldwide. Working closely with retailers and the broadcast industry, he has built the PURE brand from non-existent at the start of 2002 to its current position as #1 radio supplier to the UK, and #1 digital radio supplier in the world.

Mr Crawford represents PURE and its parent company, Imagination Technologies, on various trade bodies and has been a WorldDMB steering board member since 2004.

Tony King-Smith has more than 27 years experience in the semiconductor and consumer electronics industries. He is currently Vice President of Marketing for Imagination Technologies, where he is responsible for strategic and tactical marketing of all their semiconductor Intellectual Property (IP) technologies (including their DAB/T-DMB/DAB+ offerings). He also manages Imagination's strategic partnerships and ecosystem programmes, and sits on the Board of the company’s subsidiary, PURE Digital.

Mr King-Smith graduated in Electronics and Electrical Engineering from the University of Melbourne. He then moved to Europe, where he progressed through a diverse range of multinational companies including Panasonic, Hitachi (now Renesas), LSI Logic, INMOS and British Aerospace. Through his work, he has developed a unique global perspective of the electronics industry from European, Asian, Japanese and US management perspectives.

standard receiver chips that only decode the basic transport stream, while they re-use their multi-media processors for audio decode and data services processing.

The semiconductor progress in chipsets that are capable of demodulating and decoding DAB and DAB+ continues to advance at a dizzying pace. In addition to ever-more mature dedicated DAB solutions from companies such as Frontier Silicon, the growing market for multi-standard mobile TV receiver chips and modules supporting T-DMB is now creating increasing opportunities for adding DAB receivers to other types of consumer products.

Conclusions

DAB has had a somewhat chequered reputation around the world to date but, when examined in detail, it can be seen that it has succeeded impressively where broadcasters have recognized the critical factor – that the transition to digital radio can only happen if the consumer is presented with very clear advantages, and in particular with new and unique digital-only content.

When this happens, radio consumers – particularly in Europe where radio listening is a fundamental part of the societal psyche – are very keen to embrace the new technology, especially when dressed up in a non-threatening form. This has enabled a strong digital transition in many countries and that in turn has allowed for the development of new and very desirable features and advantages in advanced digital radio receivers.

> ***"With the advent of the DAB+ standard ... many new markets are showing keen interest to drive ahead with DAB digital radio at the heart of their radio strategy"***

With the advent of the DAB+ standard and its resulting dramatic increase in broadcast efficiency, many new markets are showing keen interest to drive ahead with DAB digital radio at the heart of their radio strategy. The first DAB+ upgradeable receivers – the PURE Siesta, Chronos II and Highway – are already shipping in volume. With many more to follow (80% of PURE's product range is scheduled to be DAB+ by the end of 2008) and DAB+ products coming from other manufacturers, we can look forward to a new era of dramatic growth in DAB digital radio around the world.

Super Hi-Vision

— research on a future ultra-HDTV system

Masayuki Sugawara
NHK

This article briefly describes the current status of R&D on the "Super Hi-Vision" television system in Japan. The R&D efforts on Super Hi-Vision are intended to explore the next-generation television system to succeed HDTV at some point in the future, and it consists of ultra-HD images and three-dimensional multichannel sound. The conceptual ideas behind the research project and the status of the technological developments are described. The collaboration between various EBU Members and NHK is also mentioned.

HDTV is now being deployed throughout the world and many broadcasters are devoting resources to HD programme production and delivery. In the meanwhile, it might be time for R&D departments to think about the future of television and broadcasting. When thinking about the future, it is always useful to look back on the past. Television history from the time of Baird and Takayanagi to the advent of SDTV was based mainly on increasing the number of scanning lines to achieve higher definition. In fact, some of the TV systems that were developed during this period were even called "high definition".

Efforts to enlarge the TV screen were accelerated after standardization of NTSC, PAL and SECAM brought the scanning-line competition to an end. Enlargement of the screen meant an enlargement of the visual field occupied by its image, that is, matching the performance of the TV to the human visual system (HVS).

Fig. 1 illustrates the changes in TV screen sizes in Japan since the 1950s. TV screens have consistently become larger, from 12 inches in the 1950s to the current 50 inches. Two difficulties arose as TV screen sizes grew. One is the picture quality degradation caused by the shortening of the relative viewing distance – since the absolute viewing distance at home does not change much. The other is how to make such a large screen. HDTV was the solution to the first problem, and flat-panel displays were the solution to the second one.

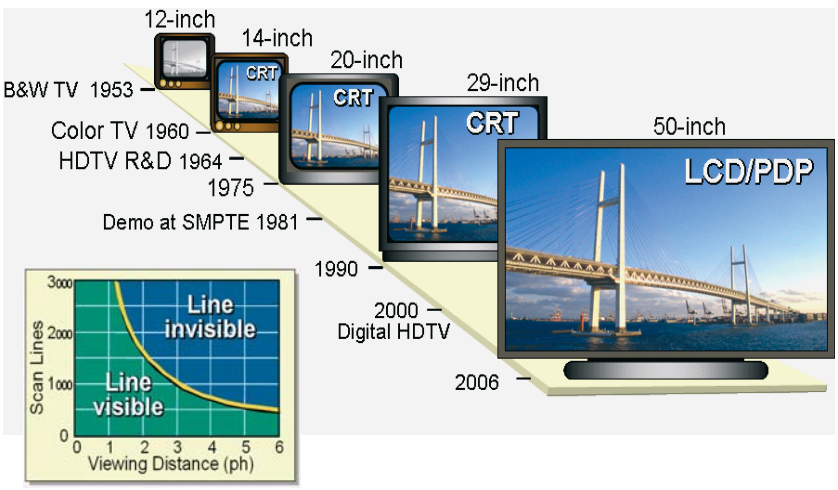


Figure 1
HDTV plays a key role in introducing a large screen into the home.

Beyond HDTV: why 8k x 4k?

HDTV was to be an almost "transparent" medium when viewed at a distance of three picture heights [1]. Its recent popularity seems to attest to this. Nevertheless, we know that HDTV is not genuinely transparent nor the final development. David Wood, Head of New Technology at the EBU, mentioned in the July 2007 issue of EBU Technical Review [2] that public expectations for quality rise over time, and are part of the process of human self-education. The more we see higher quality, the more we become accustomed to it, and the less we accept lower quality. It is human nature to seek better visual experiences, and today's HDTV might be considered just a step in the journey to match the range of the human visual system. To reach the end of this journey, quite a few parameters need to be investigated.

Among the parameters we consider most important is coverage of the visual field. To what extent should the visual field be covered by the screen? 360º surround screens and head-mounted displays with motion tracking are under study in the field of virtual reality. However, ours is a television application, and our target coverage should reflect that. Similar to the situation of viewing current TV screens, the screen would thus be rather flat, displaying a 2D image, and it would be viewed by individuals or groups.

The psychological effect of widening the visual angle usually appears as an increase in the sensation of presence or immersion in the image. These effects have been studied when designing new TV systems; in fact, they were studied in the early stages of HDTV development (e.g. [3][4]). Researchers at NHK recognized the importance of preceding studies and incorporated their findings into their plans for Super Hi-Vision. In addition, they conducted a new series of subjective and objective (physiological) experiments using Super Hi-Vision prototype equipment. These experiments assessed the relationship between the visual angle and the sensation of presence. The results indicate that the sensation of presence tends to level off at a visual angle of around 80 - 100 arc-degrees [5][6]. We consider this to be the maximum visual angle that Super Hi-Vision should provide to viewers.

The required angular resolution is needed to determine the pixel count of the image format, once the required visual angle is specified. Similar to the case of the visual angle, NHK researchers learned from the past literature (e.g. [7]) and conducted another series of experiments. These experiments used the discrimination threshold of resolution and the sense of realness as indices to determine the required angular resolution. The results were not far from what is generally believed to be necessary, i.e. 1 pixel per 1 arc-minute, or 30 cycles per degree (cpd), but a moderately higher resolution of 40~50 cpd is desirable for our criteria [8].

The primary difference between Super Hi-Vision and previous TV formats is the pixel count. However, other parameters are also important because they are influenced by the change in pixel count and/or visual angle. Also, even if they are not directly affected by the change in pixel count, these parameters need to be set to produce image quality and performance suitable for the increased pixel count of Super Hi-Vision. The frame frequency needs to be reviewed in light of both of those reasons. And while the increase in pixel count does not affect the colour space and transfer

Abbreviations

ARIB	Association of Radio Industries and Businesses (Japan)
CMOS	Complementary Metal-Oxide Semiconductor
DVA	Dynamic Visual Acuity
HVS	Human Visual System
ITU	International Telecommunication Union http://www.itu.int
ITU-R	ITU - Radiocommunication Sector http://www.itu.int/publications/sector.aspx?lang=en&sector=1
LSDI	Large Screen Digital Imagery
NTSC	National Television System Committee (USA)
PAL	Phase Alternation Line
QPSK	Quadrature (Quaternary) Phase-Shift Keying
SECAM	*Séquentiel couleur à mémoire*
SMPTE	Society of Motion Picture and Television Engineers (USA) http://www.smpte.org/
TWTA	Travelling-Wave-Tube Amplifier

function directly, it is advisable to give sufficient consideration to the view-point of pursuing higher image quality and the fact that display device technology is changing rapidly.

The current version of the Super Hi-Vision prototype is based on the conventional parameter values described in the table below but, for practical reasons, several research efforts have already started at the NHK laboratory. Of these, a recent result of a study on the relationship between dynamic visual acuity (DVA) and visual angle shows that DVA tends to increase as displays become wider in terms of visual angle [9]. This may be taken into account to determine the required frame frequency. Similarly, a review of every picture parameter would be helpful to assure a new television system will have the appropriate characteristics. The pursuit of these research activities is very desirable.

Table 1
Picture characteristics of the prototype Super Hi-Vision system

Picture aspect ratio	16:9
Horizontal pixels	7680
Vertical pixels	4320
Frame frequency	60
Image structure	Progressive
Bit/pixel	10
Colorimetry	Rec. 709

Beyond 5.1 surround sound: why 22.2?

Sound is a vital component of television systems. Since Super Hi-Vision is designed to deliver enhanced "presence", it should incorporate an audio format that matches its visual impact. Particularly, the ability to localize sound content over the widened screen images should be improved so that the horizontal and vertical image and sound match. Conventional multichannel audio systems such as the 5.1 surround sound system prioritize frontal sound reproduction at the expense of rear sound reproduction. In the case of Super Hi-Vision, it aims to provide an immersive sensation; the sound system therefore has to provide a sound field that surrounds viewers with various sound sources. Based on these requirements, the sound system on a Super Hi-Vision should:

- localize frontal sounds stably over the entire screen area;
- reproduce sound images in all directions around a viewer including elevation;
- reproduce a three-dimensional spatial impression

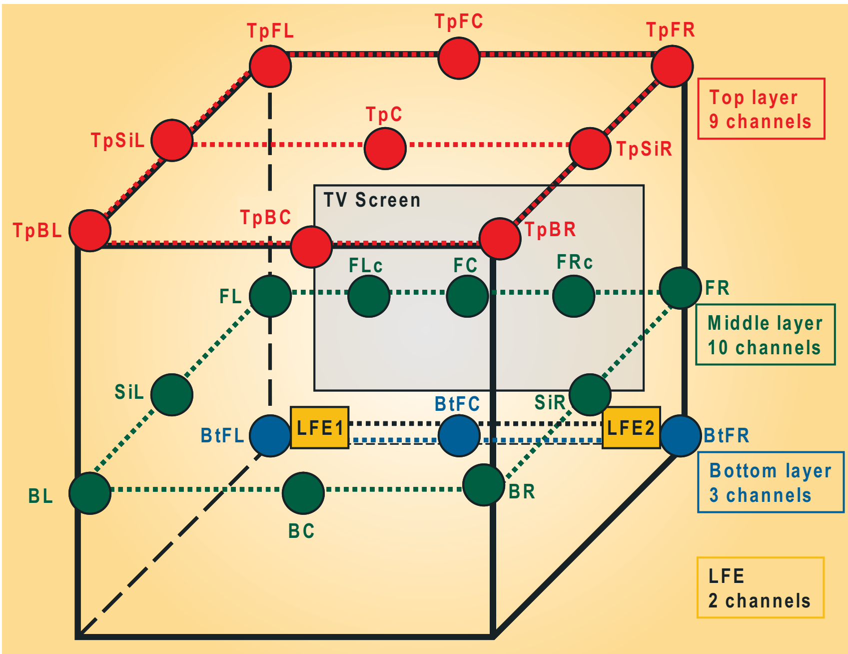


Figure 2
22.2 multichannel sound system

that augments the sense of reality;

- create a wide listening area with exceptional sound quality;
- be compatible with existing multichannel sound systems.

The 22.2 multichannel sound system was developed to fulfil these conditions. As shown in *Fig. 2*, the system consists of loudspeakers with a top layer of nine channels, a middle layer of ten channels, and a bottom layer of three regular channels and two low frequency effects (LFE) channels.

Subjective evaluations were conducted to assess the effectiveness of listening areas of three different multichannel audio systems: 2-channel stereo, 5.1 surround sound, and 22.2 multichannel sound. The results indicate that three-dimensional sound by a 22.2 multichannel sound system can produce better sensations of spatial sound quality, reality and presence in a wider listening area than 5.1 and 2.0 multichannel sound systems [10].

Status of R&D

Fig. 3 shows the Super Hi-Vision devices developed so far and the system configuration.

The first demonstration was in 2002, and it featured a prototype consisting of a camera, projection display, and frame grabber. A newer prototype was exhibited for six months at World Expo 2005 in Aichi, Japan, where it showed special Super Hi-Vision programmes to approximately 1.56 million people. The prototype system was also installed at a museum. Since 2006, it has been demonstrated in the USA (NAB 2006 and 2007) and in Europe (IBC 2006). During this time, work has continued on various elemental technologies, such as optical transmission of uncompressed video and MPEG-2 compression devices. Experiments have been conducted on indoor transmission in

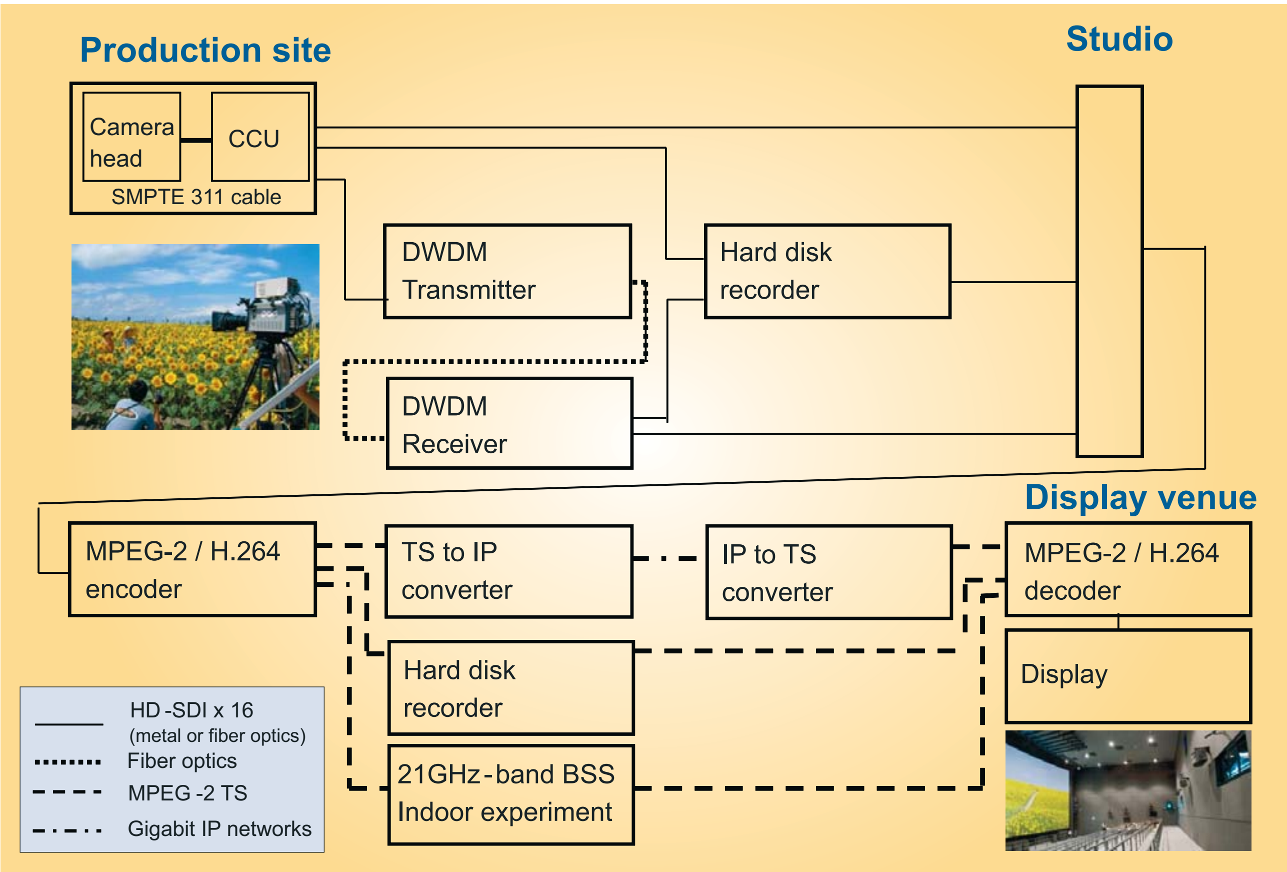


Figure 3
Current system configuration of Super Hi-Vision

the 21 GHz band and on real-time transmission of live TV programmes using the IP transport protocol.

Currently, research at NHK concentrates on developing image and display devices with the "full" resolution, i.e. 33 million pixels. Such technologies were not available at the beginning of this project; the total resolution of the prototypes had to be enhanced by pixel-offset imaging.

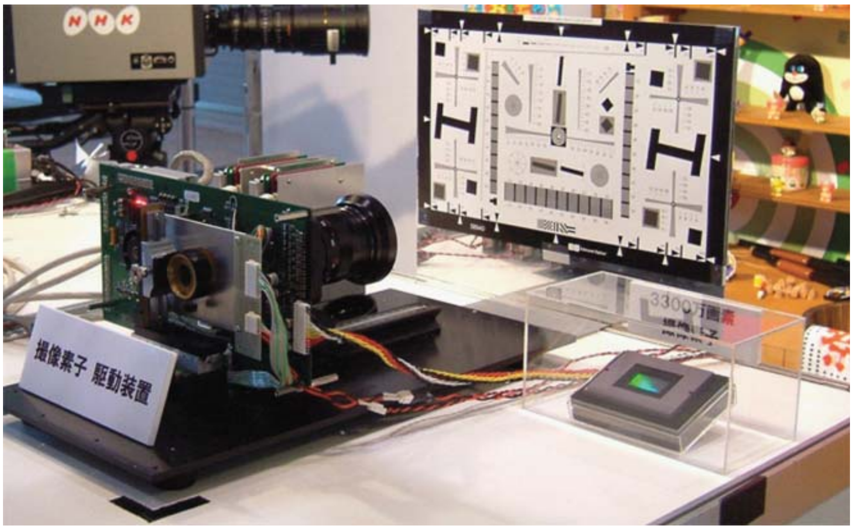


Figure 4
33 megapixel image sensor for Super Hi-Vision demonstrated at NHK Science and Technical Laboratories open house in May 2007

A full-resolution image sensor was demonstrated in May 2007 at the annual open house of the NHK Science and Technical Research Laboratories *(Fig. 4)*. The newly-developed CMOS image sensor has the pixel count of 7680 x 4320 pixels with 60 frames per second output. The advent of a "full-resolution" camera and display will enable the true picture quality of Super Hi-Vision to be demonstrated.

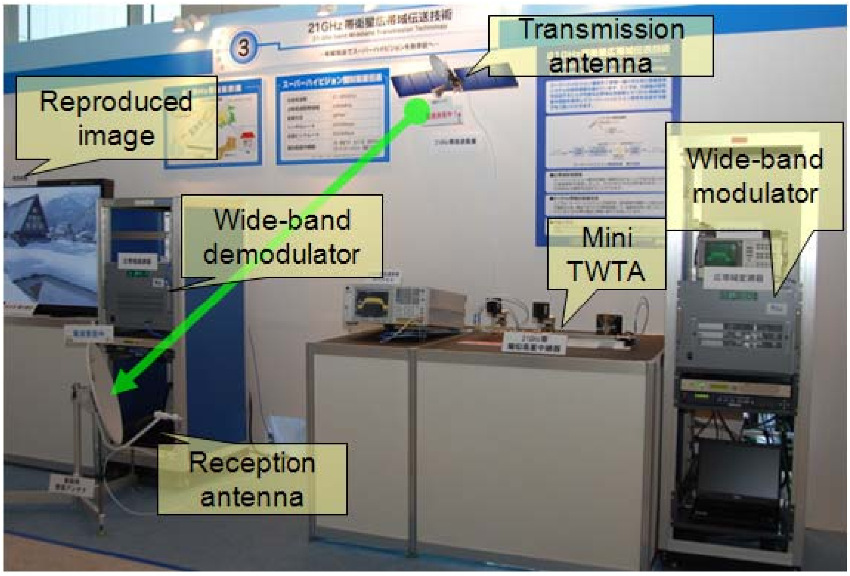


Figure 5
21-GHz-band indoor transmission experiment at open house in May 2007

The means to deliver programmes to the home and display them are big issues. Besides a broadband communications network, a broadcasting satellite system operating in the 21 GHz band is a promising candidate for the transmission path to the home. The bandwidth of the channel assigned to the 21 GHz band is 600 MHz and therefore it has the potential to handle extremely high bitrate signals. An indoor transmission experiment was demonstrated at the NHK open house in 2007 *(Fig. 5)*.

In this demonstration, a Super Hi-Vision signal compressed by MPEG-2 at 250 Mbit/s was input to a wideband modulator (300 MHz bandwidth) that can transmit a 500 Mbit/s signal by using QPSK modulation. The mini travelling wave tube amplifier (TWTA) performed wideband amplification of the signal before it was sent to the receiver on the floor via the dummy satellite antenna on the wall. The transmission length of this experiment was only two meters but it did show the feasibility of transmitting a Super Hi-Vision signal from a satellite located 36,000 km away from the earth to the home.

Needless to say, standardization of television technologies is very important. Regarding image formats, the three standards listed below deal with the image format of Super Hi-Vision.

- ITU-R Recommendation BT.1201 (1995-2004): **Extremely high resolution imagery**

Masayuki Sugawara received B.Sc. and M.Sc. degrees in electric communication engineering, and a Ph.D. degree in electronic engineering from Tohoku University, Sendai, Japan. He joined NHK (Japan Broadcasting Corporation), Tokyo, in 1983. Since 1987, he has been researching solid-state image sensors and HDTV cameras at NHK's Science and Technical Research Laboratories (STRL).

Dr Sugawara was an associate professor at the University of Electro-Communications in Tokyo from 2000 until 2004. At present, he is a senior research engineer at NHK STRL and is engaged in the research of ultra-high-definition TV systems. He is a member of SMPTE, IEEE, IEICE and ITE.

- ITU-R Recommendation BT.1769 (2006): **Parameter values for an expanded hierarchy of LSDI image formats for production and international programme exchange**
- SMPTE 2036 (2007): **Ultra high definition television – Image parameter values for program production**

In addition to these standards, standardization activities are ongoing in several bodies, including the ITU, SMPTE, and ARIB.

EBU and NHK collaboration on the future of television

Looking back at the history of HDTV, we can see that it takes tremendous effort to develop a new medium and deploy it. The same can be said of Super Hi-Vision or any kind of new media that goes beyond HDTV, even though the speed of technology advancement is accelerating. It would be impossible for any organization to finish such a difficult project on its own. Inter/intra industry collaboration would be desirable because of the public characteristics of broadcasting and programme exchanges. In this respect, collaboration between broadcasters is of the highest priority.

The BBC, RAI, IRT (all EBU Members) and NHK agreed to collaborate on broadcast technology R&D in February 2007. Moreover, in February 2008, they agreed to conduct joint research on the items listed below.

- High-efficiency image coding technology for Super Hi-Vision and virtual studio technology (BBC and NHK);
- broadcasting satellite transmission technology in the 21 GHz band (RAI and NHK).

Along with these joint research projects, the collaborators carry out discussions and information exchanges on several topics, including technologies for assisting disabled viewers.

Broadcasters are now facing harder competition than ever for the attention of viewers at home. Therefore, it is a wise move to tackle the technological developments in a cooperative way. We, at NHK, believe that this collaborative effort will bear much fruit.

References

[1] ITU-R Report BT.801: **The Present State of High-definition Television**

[2] David Wood: **The development of HDTV in Europe – a tale of three cities: Dublin, Dubrovnik and Geneva**
EBU Technical Review, July 2007.

[3] T. Hatada, H. Sakata and H. Kusaka: **Psychophysical Analysis of the "Sensation of Reality" Induced by a Visual Wide-Field Display**
SMPTE Journal, vol. 89, pp. 560-569, 1980.

[4] N. E. Tanton and M. A. Stone: **HDTV DISPLAY: Subjective effects of scanning standards and domestic picture sizes**
BBC RD 1989/9.

[5] K. Masaoka, M. Emoto, M. Sugawara and Y. Nojiri: **Contrast effect in evaluating the sense of presence for wide displays**
Journal of the SID, vol. 14, pp. 785-791, 2006.

[6] M. Emoto, K. Masaoka, M. Sugawara and F. Okano: **Viewing angle effects from wide field video projection images on the human equilibrium**
Displays, vol. 26, pp. 9-14, 2005.

[7] J. O. Drewery and R. A. Salmon: **Tests of visual acuity to determine the resolution required of a television transmission system**
BBC R&D White Paper, WHP 092, 2004.

[8] K. Masaoka, T. Niida, M. Murakami, K. Suzuki, M. Sugawara and Y. Nojiri: **Perceptual limit to display resolution of images as per visual acuity**
Presented at Electronic Imaging 2008.

[9] Y. Matsuo, M. Emoto and M. Sugawara: **Frame Frequency for Wide Field of View Video System – Dependency of Field of View on Dynamic Visual Acuity**
IEICE Forum on Information Technology, Toyota, 2007.

[10] K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama and A. Ando: **A 22.2 Mutlichannel Sound System for Ultrahigh-Definition TV (UHDTV)**
SMPTE Motion Imaging Journal, vol. 117, no. 3, pp. 40-49, 2008.

SVC — a highly-scalable version of H.264/AVC

Adi Kouadio
EBU Technical

Maryline Clare and Ludovic Noblet
Orange Labs, France Telecom R&D

Vincent Bottreau
Thomson Corporate Research

Scalable Video Coding (SVC) is a recent amendment to the ISO/ITU Advanced Video Coding (H.264/AVC) standard, which provides optional but efficient scalability functionalities on top of the high coding efficiency of H.264/AVC. In addition to bringing a cost-efficient solution to the delivery of different formats of the same content to multiple users, it can be used to provide a better viewing experience (enhanced content portability, device power / content-quality adaptation, fast zapping times and fluid forward / rewind functions, efficient error retransmission, etc.).

This article describes the potential of SVC, in terms of applications and performance. A brief overview of SVC functionalities, as well as practical use cases, are given in the following sections. Different performance evaluations, based on test results, are also described.

1. Introduction

Providing content everywhere is a major goal for video service providers. In addition to legacy broadcast TV, consumer video applications today span:

- IPTV (over managed networks and the open Internet);
- catch-up TV;
- Video-on-Demand services (VoD);
- Mobile TV;
- Web 2.0 content (including user-generated) and media platforms, etc.

All these new video applications are becoming a reality, thanks to developments in transmission, storage and compression technologies. Another enabler for this diversity of services is the strong penetration of end-user devices such as HDTV flat-panel displays, portable multimedia players (PMPs) and 3G mobile devices – and the availability of broadband Internet access ... providing high bandwidth connectivity into the home (xDSL, FTTH), within the home, between devices (Ethernet, Wi-Fi, Power-Line Communication) and outside of the home (3G, 4G, WiMax).

Nevertheless, providing content everywhere in such an environment, while achieving cost efficiency, is still a challenge for video service providers. Enabling such services implies the implementation of content-repurposing bricks in the service architecture, for transcoding between:

- *multiple image formats* (QCIF, CIF, QVGA, VGA, SD, HD);
- *multiple bitrates* (variable or constant according to the access networks);
- *multiple frame rates* (50 Hz, 25 Hz, 12.5 Hz) and;
- *different delivery platforms (with different coding schemes)*.

Repurposing (transcoding) the content at any point in the chain generates extra cost, either for the service provider, the consumer or the network operator. It also alters the user experience by being an additional hurdle to content portability, which does not necessarily preserve the included DRM (Digital Rights Management). Other alternatives, such as content simulcasting, would result in higher bitrate requirements.

Fortunately, while many different video codecs were used in the past (depending on the targeted device for a given service), there is today a general trend towards H264/AVC [1]. This codec has not only been widely implemented in set-top boxes, it is also soon to be generalized in mobile devices as well as in Portable Media Players (PMPs); it has even been introduced recently in the Adobe Flash 9 and Apple QuickTime media players.

With H264/AVC generalization, transcoding will soon be limited to format and bitrate adaptation, and there will be less need for several output codecs. This is a key point in considering the SVC scalable extension of H264/AVC [2].

Considering the continuously increasing number of possible combinations between formats and bit-rates, smart content adaptation today becomes a key issue for achieving the “content everywhere” target.

Because of all these considerations, scalability and flexibility are key points for the near future of video services, whether these are new services or the evolution of existing services. Such scala-bility is needed not only at the architecture and infrastructure levels, but also at the content level.

Scalable Video Coding provides the appropriate tools to efficiently implement content scalability and portability. It is the latest scalable video-coding solution, and has been standardized recently as an amendment to the now well-known and widespread H.264/AVC standard [1] by the *Joint Video Team* [1] (JVT). Other video scalability techniques have been proposed in the past, (even standard-ized as optional modes for MPEG-2 [3] and MPEG-4 - Part 2 [4]) but they were less efficient and more complex; moreover, because of the (then) lack of market need for scalability (video services limited to standard-definition broadcasts), they were never used.

In the following sections, we will first give a brief technical overview of SVC functionalities. The second part will outline different practical use-cases of the standard while the third part will describe preliminary performance evaluation results. The fourth and last part will describe ongoing work and the action taken by the EBU and other standardization bodies (MPEG, JVT) to extend the standard and provide more clarity on SVC performances.

2. SVC overview

Scalable coding consists of compressing a digital video into a single bitstream in such a way that other meaningful and consistent streams can be generated by discarding parts of the original

1. The Joint Video Team is a joint working group from the ITU VCEG group and the ISO/IEC MPEG group at the origin of the H.264/AVC standard.

compressed stream. Those sub-streams can be directly interpreted at different bitrates, different resolutions or different time scales.

SVC organizes the compressed file into a **base layer** (BL) that is H264/AVC encoded, and **enhancement layers** (EL) that bring additional information about quality, resolution or frame rate *(see Fig. 1)*. This implies that SVC base-layer streams can be decoded by H.264/AVC products (set-top boxes, PMPs), thus ensuring backward compatibility for consumers not having the SVC upgrade. More information about H264/AVC can be found in [1].

Figure 1
Overview of SVC layer structure (EL = enhancement layer)

SVC provides spatial, quality and temporal scalability types *(see Section 2.1)* that can be combined at each level. The enhancement layers can be fully hierarchical, or not:

- a layer can be generated ("predicted") from another layer and also be a prediction for yet another one;
- or a layer can be a prediction for two layers that are not hierarchically inter-dependent.

Such encoding parameters depend on the targeted application.

A compressed video bitstream is made up of Network Abstraction Layer (NAL) units, and each enhancement layer corresponds to a set of identified NAL units. A NAL unit is a packet with a header of a few bytes (containing information about the payload) and a payload corresponding to the compressed information. A set of successive NAL units, sharing the same properties, forms a NAL access unit.

Depending on the context, the enhancement layers may (or may not) be transmitted by the network, and may (or may not) be decoded by the end user device. In the first case, the network integrates some adaptation modules – deciding what to transmit and what to filter (for instance, depending on the network bandwidth characteristics). In the latter case, the terminal extracts the layers it can exploit. Such an adaptation mechanism is based on packet selection / dropping.

Setting up a service based on SVC technology implies two important considerations, ***decision*** and ***adaptation***, which are further discussed in *Section 2.3*.

2.1. Scalability

2.1.1. Spatial scalability

Spatial resolution gives the video horizontal and vertical dimensions in pixels, resulting in several well-known "video formats" such as QCIF (176 x 144 pixels), CIF (352 x 288), SD (720 x 576) and HD (1280 x 720 up to 1920 x 1080).

The SVC standard's ability to embed 4:3 and 16:9 picture aspect ratios is, for example, a very important spatial scalability feature, typically when considering SD/HD broadcast. It should be noted that, depending on the standard profile in use, the ratio between layers can be fixed to a restricted set of values *(see Section 2.2)*.

Spatial scalability is provided by filtering / upsampling mechanisms and inter-layer predictions (Motion data prediction, intra-texture prediction and residual signal prediction). Each spatial enhancement layer (EL) is referred to as a *dependency representation*. A predicted EL always indicates the reference layer representation where it was originally predicted. EL macroblocks are predicted from reference layer macroblocks. They inherit motion vector values and other prediction data (texture and residual) from the appropriate reference layer macroblocks, after normative scaling and merging processes.

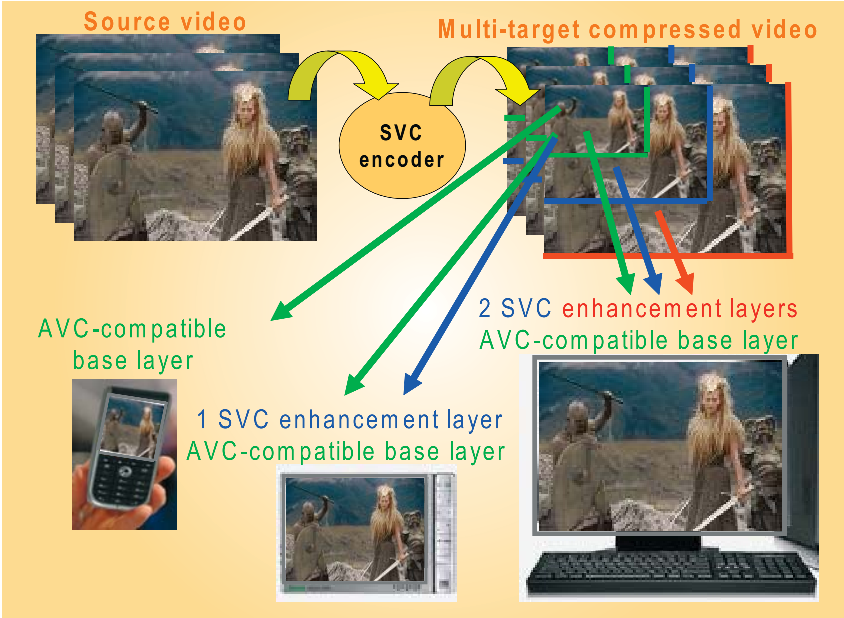


Figure 2
Spatial scalability

Spatial scalability can typically be used for transmission of the same video bitstream to PCs and portable devices *(see Fig. 2)*, or to SD and HD television sets.

2.1.2. Temporal scalability

Temporal scalability defines the difference in the number of images per second (expressed in Hz). Typical frequencies in Europe are 50 Hz, 25 Hz or 12.5 Hz.

SVC extends the tools already provided by H264/AVC (hierarchical P or B slices coding), structuring the bitstream into a hierarchy of images, thus allowing the easy removal of the lower level(s) in the hierarchical description.

Temporal scalability can typically be used in video transmissions over mobile networks where bandwidth capacity can change very often, or if the target terminal has very low CPU capacities: in such cases, it is interesting to drop the enhancement layers and send only the base layer (which could, for example, contain only half the number of images per second).

2.1.3. Quality scalability

Quality scalability is often referred to as SNR (Signal-to-Noise Ratio) scalability and is intended to give different levels of detail and fidelity to the original video, while having the same spatial and temporal definitions. In an SVC-compressed bitstream, each spatio-temporal layer can have different levels of quality – each of them bringing additional detail and accuracy.

It is up to the encoding process to decide whether more detail will be added to random parts of the video images, or to some specific parts of given images. Differences in quality levels can be medium (MGS, Medium Grain Scalability) or large (CGS, Coarse Grain Scalability). CGS provides a quality difference of about 25% between two layers while MGS offers 10%. MGS uses a modified high-level signalling, which allows a switching between different MGS layers in any access unit, and the so-called key picture concept, which allows the adjustment of a suitable trade-off between drift and enhancement layer coding efficiency for hierarchical prediction structures.

With the MGS concept, any enhancement layer NAL unit can be discarded from a quality-scalable bitstream, and thus packet-based quality-scalable coding is provided. The possibility of very fine granularity (FGS, Fine Grain Scalability), resulting in bitstreams that can be truncated anywhere, has been considered during the joint MPEG/ITU standardization process, but was not finally selected – as the finer the quality scalability is, the more complex the encoding/decoding process. Alternatively, the MGS concept allows the EL transform coefficients to be distributed among several slices. Thus, the information for a quality refinement picture that corresponds to a certain quantization step size can be distributed over several NAL units corresponding to different quality refinement layers.

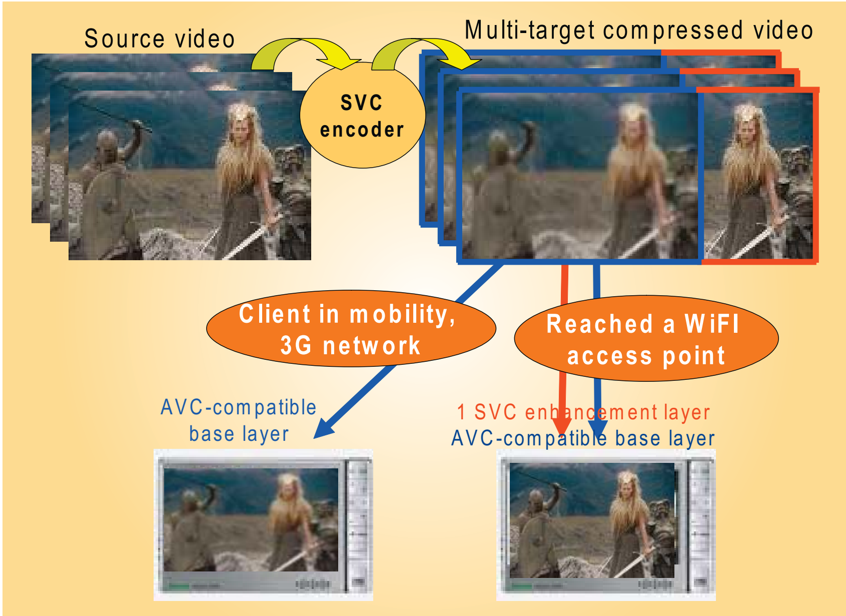


Figure 3
Quality scalability

Quality scalability can typically be used for:

- HD transmission to customers that are eligible for HD (full quality) and people not eligible for HD quality, but still equipped with HD screens (top enhancement layer is dropped);
- or for extra refinements when the bandwidth increases in mobile environments *(see Fig. 3)*.

2.1.4. Interlaced and progressive scalability

SVC inherits interlace tools from the H.264/AVC system: **Paff** (Picture adaptive frame/field) and **Mbaff** (Macroblock adaptive frame/field). SVC specifies four types of scanning-mode scalability – with some subjected to coding restrictions, depending on the interlaced coding mode of the base and enhancement layers:

- progressive to interlaced;
- interlaced to interlaced;
- interlaced to progressive;
- progressive to progressive.

More information on scanning-mode scalability can be found in [5].

2.2. Profiles and Levels

Profiles define the set of coding tools (for example, arithmetic or run length entropy coding, etc.)

that can be used to build up the stream, while ***levels*** specify the constraints on key coding parameters, such as the number of macroblocks, the bitrate.

The SVC extension specifies three new scalable profiles, which are closely related to H.264/AVC profiles:

- Scalable Baseline Profile: aimed at low complexity applications;
- Scalable High Profile: aimed at broadcasting and video storage applications;
- Scalable High Intra Profile: aimed at professional applications.

Table 1
SVC profiles – main differences

Tools\Profiles	Scalable Baseline	Scalable High	Scalable High Intra
Base Layer profile	AVC baseline	AVC High	AVC High intra
CABAC and 8x8 transform	Only for certain levels	Yes	Yes
Layer spatial ratio	1:1 , 1.5:1 or 2:1	Unrestricted	Unrestricted
I, P and B slices	Limited B slices	All	Only I slices
Interlaced Tools (Mbaff & Paff)	No	Yes	Yes

The levels are the same as the H.264/AVC levels. However, the characteristic number of macroblocks per second in an SVC stream is calculated according to the number of layers in the stream *(see formula below)*.

$$N = \begin{cases} \frac{1}{2} * \sum_{i=1}^{L-2} NL_i + NL_L, & L > 2 \\ NL_2 \quad , & L = 2 \end{cases} \quad \text{where } L > 0$$

with,
N = Total number of Macroblocks.
L = Total number of layers in the stream.
NL_i = Number of Macroblocks on layer i

2.3. Overview of a service architecture with SVC

Once an SVC bitstream is generated, its scalability properties enable it to best match the transmission conditions over a network path to a given location and end-user device. One or more adaptation mechanisms can be implemented somewhere in the content delivery network – between the video streaming source and the end-user device. Such an adaptation process needs to obey a decision mechanism, based on the complete context at the time the service is offered (terminal properties, subscription characteristics, available bandwidth, error rate, DRM, etc.). Typically, in an IMS/TISPAN environment, it shall be noticed that previously-mentioned contextual information can be processed in very close relationship with not only access-control-related functions of IMS/TISPAN, but also with user-descriptive data.

Depending on the application, the decision and adaptation processes might not be implemented in the same equipment. Data flows in decision and adaptation mechanisms are illustrated in *Fig. 4*. Such a decision can be static (made once only, at the start of the transmission) or dynamic, and can be implemented in different parts of the service architecture, for example:

- At encoding – by splitting the video units into different bitrate levels, all assigned a different output stream, and sent to user groups having heterogeneous network / terminal capacities;
- At a VoD server – by adapting the video transmission bitrate in a unicast session, or by using scalability properties for providing a better trick mode user experience;
- At a network node – by reorganizing the video units to allow video streaming over a sub-network with different characteristics;

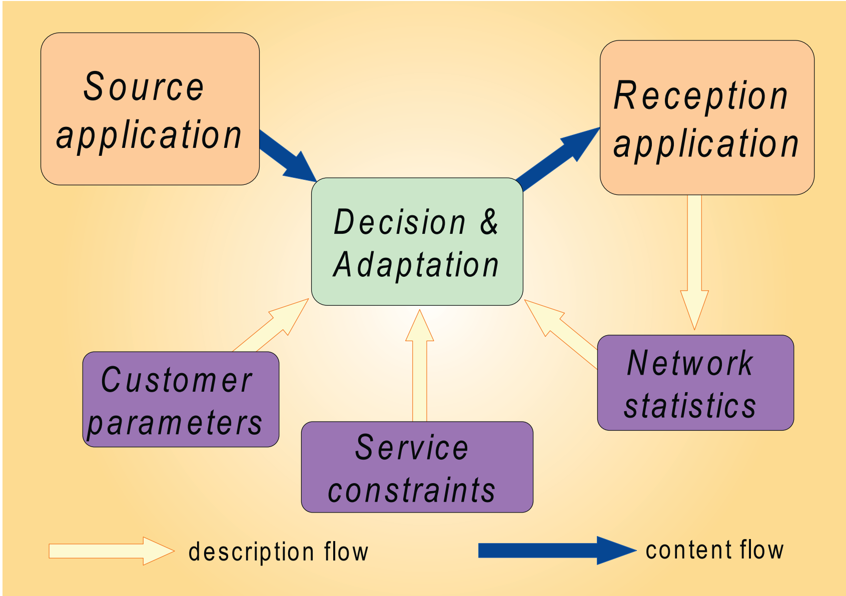


Figure 4
Bitstream adaptation

Abbreviations

3GPP	3rd Generation Partnership Project
ATSC	Advanced Television Systems Committee (USA)
AVC	(MPEG-4) Advanced Video Coding, part 10 (aka H.264)
BL	Base Layer
CABAC	Context-Adaptive Binary Arithmetic Coding
CAS	Conditional Access System
CDN	Content Delivery Network
CGS	Coarse Grain Scalability
CIF	Common Interchange Format (352x288 pixels)
DLNA	Digital Living Network Alliance http://www.dlna.org/home
DMB	Digital Multimedia Broadcasting http://www.t-dmb.org/
DRM	Digital Rights Management
DSLAM	Digital Subscriber Line Access Multiplexer
DSP	Digital Signal Processor / Processing
DVB	Digital Video Broadcasting http://www.dvb.org/
EL	Enhancement Layer
FGS	Fine Grain Scalability
FTTH	Fibre To The Home
GoP	Group of Pictures
HDTV	High-Definition Television
HHI	Heinrich Hertz Institut (German R&D lab)
IC	Integrated Circuit
IMS	IP Multimedia Subsystem
IPTV	Internet Protocol Televison
JSVM	Joint Software Verification Model
JVT	(MPEG/VCEG) Joint Video Team
LCD	Liquid Crystal Display
Mbaff	Macroblock adaptive frame/field
MGS	Medium Grain Scalability
MOS	Mean Opinion Score
NAB	National Association of Broadcasters (USA) http://www.nab.org
NAL	Network Abstraction Layer
Paff	Picture adaptive frame/field
PLC	Power-Line Communication, also written PLT, BPL ...
PLT	Power-Line Transmission/Telecommunication, also written PLC, BPL ...
PMP	Portable Multimedia Player
PoP	Point Of Presence
PSNR	Peak Signal-to-Noise Ratio
QCIF	Quarter Common Intermediate Format (176x144 pixels)
RWTH	*Rheinisch-Westfälische Technische Hochschule*
SAMVIQ	Subjective Assessment Methodology for Video Quality
SDTV	Standard-Definition Television
SNR	Signal-to-Noise Ratio
SSIM	Structural Similarity Metric
STB	Set-Top Box
SVC	(MPEG-4) Scalable Video Coding
SWOT	Strengths, Weaknesses, Opportunities, Threats
TISPAN	Telecoms & Internet converged Services and Protocols for Advanced Networks
WG-IPTV	Walled Garden IPTV

- At the edge – where a DSLAM can dynamically select video units (i.e. packets) for QoS, channel change, or just eligibility management;
- At the home gateway – for adaptively taking into account home networking conditions.

3. Envisaged use cases

As mentioned earlier, the current audiovisual landscape and its permanent evolution generate a strong need for scalability. We have already identified a few use cases where SVC would be of some benefit; now we present a few that are consensually considered as potentially providing much benefit.

3.1. *Quality of Service and enhanced user experience*

3.1.1. Fast zapping, fluid forward and rewind

The goal here is to improve the customer experience when using functions such as channel change or video fast forward / rewind. If the current video is displayed at a given bitrate and is decomposed on, at least, a base layer and an enhancement layer, then switching to the lowest layer allows us to either quickly visualize the next channel (TV channel switch) or have a more fluid fast forward mechanism (VoD function).

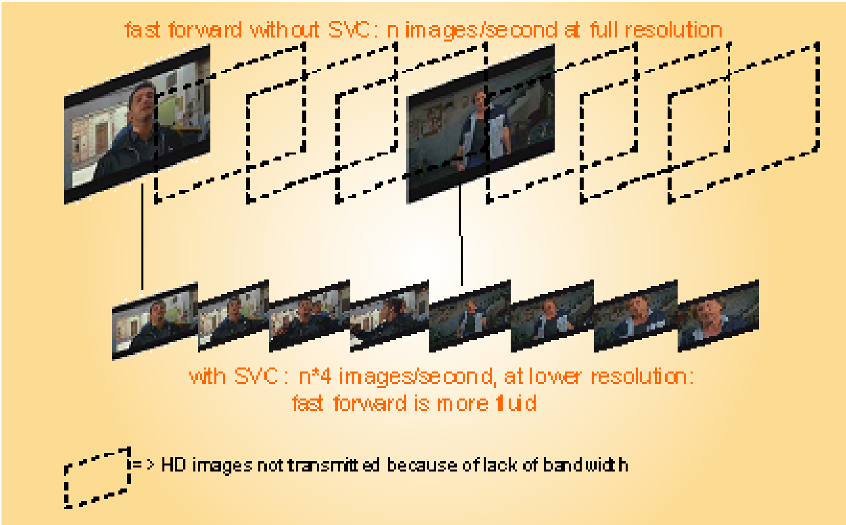


Figure 5
SVC for faster forward at lower resolution

This is explained by the fact that less information is needed to describe lower-layer images, so the available bandwidth is used to transmit more, but smaller, images *(see Fig. 5)*. Of course, the new decoded stream (a new channel in the case of a switch, or further sequences of the video in the case of a fast forward) offers a lower quality until both layers can finally be sent ... lower resolution images are, in a first phase, received and enlarged artificially to fit to the screen dimensions. It is, however, demonstrated that the human visual system needs a couple of seconds before it is really sensitive to quality, so having a black image (channel switch) or a non-fluid forward is more annoying than quickly seeing information – even if this leads to a poor quality image for up to two seconds.

3.1.2. Integration of retransmissions at constant bitrate

In the case of frequent transmission errors, different mechanisms are set up to correct them and improve the quality of service. One of these mechanisms consists of retransmitting those packets that never reached the terminal side. However, such mechanisms either require extra bandwidth or they slow down the transmission rate because of the information overhead.

An interesting feature of SVC is the ability to use the enhancement layer as a retransmission layer: in the case of errors, retransmission packets will be inserted in such layers, thus providing error correction at constant bitrate *(see Fig. 6)*. The only price to pay is to accept a loss of quality since

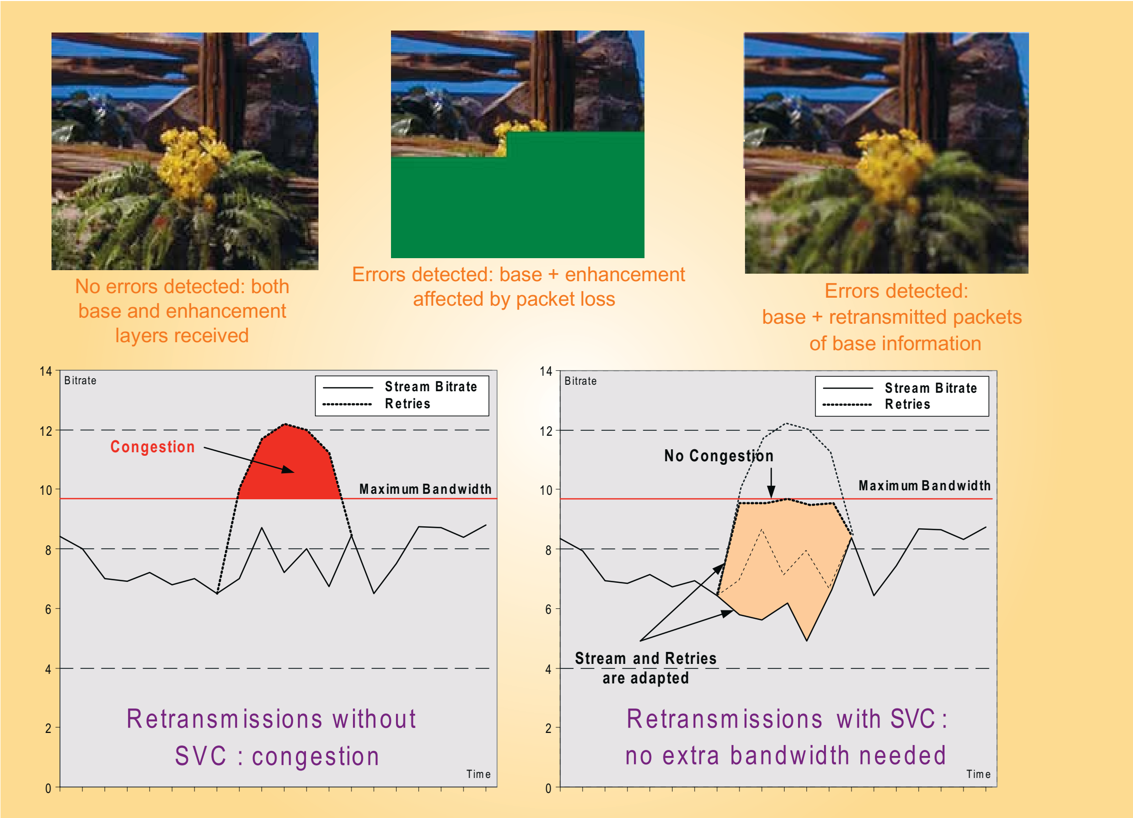


Figure 6
SVC for integrating retransmissions at constant bit-rate

less enhancement data is then received – some enhancement data being replaced by re-sent base-layer data. Such a mechanism guarantees reception of a minimum level of information quality, by providing maximal protection to this information.

3.2. Continuity of service in mobile environment

Mobile reception cannot rely on a stable bandwidth (it can drop when the network cell is overloaded or when the customer experiences a hand-over), but customers are highly annoyed by service ruptures. The advantage of having a video decomposed into layers is that they can simply be dropped and later reinserted, depending on the available bandwidth *(see Fig. 7).*

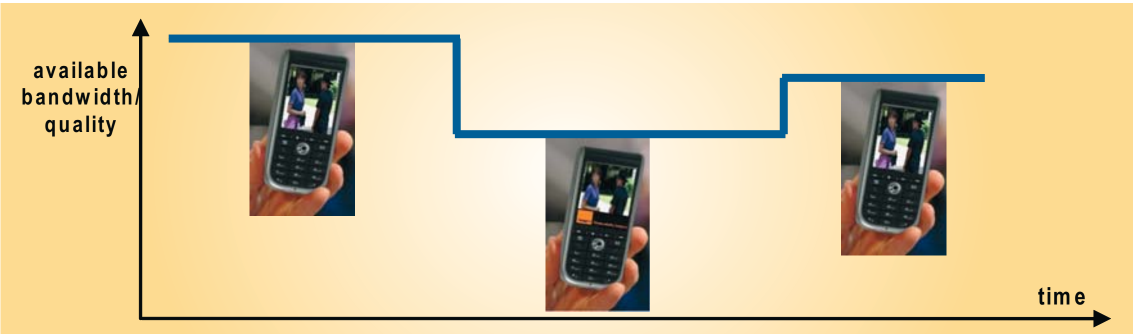


Figure 7
SVC for maintaining service in mobility

Such capacity is not only used in the case of difficult transmission over a single network, it can also be used for the same video transmitted through different types of networks and towards different types of terminals: the same video is sent with a base layer only to mobile phones, but PCs connected via 3G cards to the same network can receive the full video quality (base + enhancement).

3.3. Portability for Video on Demand: Open Internet illustration

SVC offers an extremely simple way of using the "same video" source on different terminals. No operation such as transcoding is necessary, so there is no extra computing time and no loss of quality. This becomes extremely useful when you don't know in advance where and how you will finally use the video content you are interested in. SVC allows you to buy the content, start watching it on a given terminal and finish watching it on another. A very important consequence for content owners is the fact that DRM is preserved, which is an essential advantage when compared to transcoding.

An immediate illustration of this characteristic is the naturally-convergent video platform called the Internet ! You can access its portals from PCs, but also now (and soon even more) from mobile phones. As shown in *Fig. 8*, this results in an obvious need to watch given content on a platform that is not decided in advance (i.e. at downloading time).

Figure 8
SVC for easy content portability management

3.4. Cost optimization for on-demand long tail management

As on-demand catalogues become larger and larger in terms of available titles / references, the costs for preparing and repurposing such content become higher and higher, even when using automated workflows. Moreover, this cost inflation is also related to the target devices and bitrates (access networks).

While there are known solutions to the cost model for the most popular content, the question of efficiently amortizing the preparation costs of the "long tail" (less popular content) still remains, even though the associated owners' rights are less important than for block-busters.

The above-mentioned costs correspond to the following tasks, needed for the general process of providing on-demand content:

- content capture (from tape, turn-around, file, etc.) and editing may require indexation means as well as metadata extraction / generation;
- content preparation (encoding and transcoding) and metadata processing;
- content integrity checking;
- content packaging (including protection);
- content provisioning;
- content ingest.

Content preparation and checking are important costs within this process. Multiplying the instances to be generated, in order to address multiple devices through multiple access networks, increases the overall on-demand processing costs, especially for the long-tail content – even with automated workflows.

We believe that using SVC instead of multiple H.264/AVC instances can reduce the capital and operational expenditures for the aforementioned content processing.

Moreover, we believe that SVC also allows more efficient schemes for content management within a content delivery network (CDN). Actually, it allows us to use access-network characteristics to manage the content distribution down to the edge. In that context, the combination of spatial (multiple devices) and SNR scalabilities (access-network capabilities) are probably the most anticipated features.

Finally, storage costs are lowered because the SVC global version is leaner than the cumulated AVC files of embedded versions.

3.5. *Fine adaptation to xDSL residential eligibility*

3.5.1. *Optimal quality for a given eligibility level: illustration with HD screen without HD eligibility*

SVC helps here to define intermediate ADSL eligibility levels, to which specific quality can be offered. This can be used, for instance, to provide good SD programme quality to customers who cannot reach HD bitrates, but who can still have a much better picture quality than the one provided by the SD eligibility threshold.

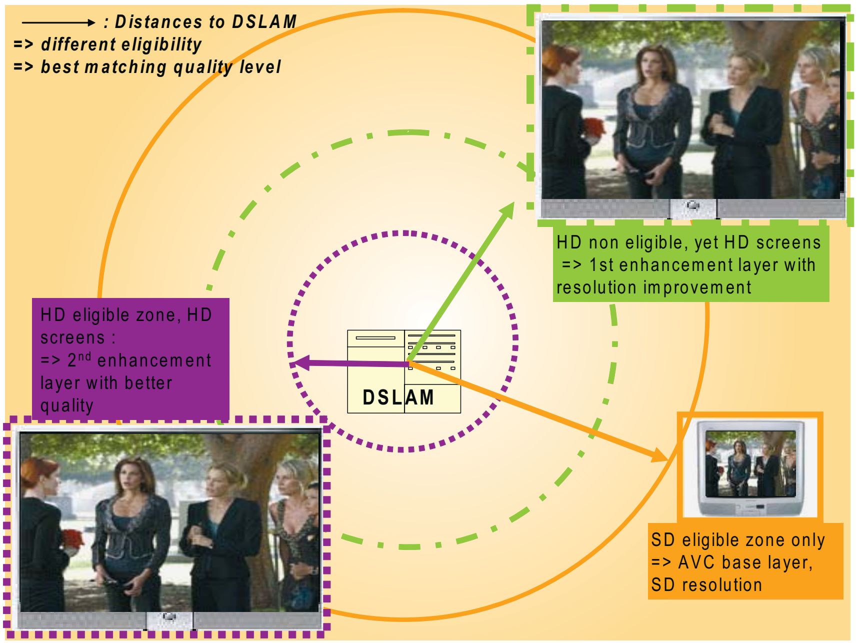


Figure 9
SVC for fine eligibility level management

The programme can be SVC-encoded, with an H264/AVC SD base layer, plus a first quality and / or resolution enhancement layer – at a bitrate that is halfway between SD and HD bitrates – and a second enhancement layer for HD customers *(see Fig. 9)*. This would help improve the bad quality noticed by customers viewing programmes on their HD screens if the quality of the video they receive is not really HD.

Such a feature can also apply to FTTH deployments. There will indeed be a period of time when FTTH is not available everywhere. SVC can provide a natural way to deliver the same content at different bitrates, qualities and resolutions, so when FTTH is present, an ultimate enhancement layer can be dedicated to the FTTH bandwidth capacities, thus allowing premium HD content to homes that can receive it.

3.5.2. *Mutual dynamic access to bandwidth inside a home*

SVC SNR scalability provides an efficient means for achieving a better simultaneous Walled Garden-IPTV (WG-IPTV) and Internet experience for the end user by adapting the Walled Garden IPTV video bitrate, depending on the bandwidth necessary for achieving the correct user experience on the Internet side within reasonable limits. Moreover, if the internet video is encoded with SVC SNR scalability, it can also be adapted so that the impact over the WG-IPTV stream is not perceptible (including with fade-in/fade-out mechanisms at the transitions). This adaptation can be either dynamically managed on the network side, or under full user control. With single-layer video technologies, the implementation of such dynamic mechanisms would require a transrating operation to be performed somewhere – and probably at the edge – which obviously does not seem to be interesting in terms of network architecture and density for massive processing and cost-efficiency.

An extension of this use case is bandwidth dynamic adaptation, for example if a second TV set (SDTV) is turned on when another (HDTV) is already being used. In this case, bandwidth can be shared between both programmes, providing optimal trade-off in quality to each, depending on their properties and on the target terminal characteristics.

3.5.3. Vector for efficient Premium HDTV Services enabling

The consumer adoption of H.264/AVC set-top boxes is still in progress. Bringing a completely new scalable system to market would imply a new set of products and interoperability issues with existing systems. Simulcasting several streams would require more bandwidth than using SVC on its own and would also require the user to tune in to the channel that provides the required quality/resolution.

SVC base layer compatibility with H.264/AVC would enable consumers who do not have SVC functionality to still see their usual programmes (SDTV or HDTV [1080i/25 or 720p/50]) if the latter is provided as the base layer of the SVC stream. SVC-compatible user will be able to access, for example, higher quality (1080p/50) signals stored in the enhancement layer as a premium service.

It is worth mentioning that the support of interlaced/progressive is very important for the broadcasting industry in order to enable a smooth (for consumers) and efficient (for broadcasters) transition from an SD to an HD (hopefully completely progressive) audiovisual landscape.

3.6. Other use cases

Even though they appeared to us as being of lower priority at the moment, other identified use cases should also be further investigated.

- Premium content incentive (“teasing”): show free content that is missing essential information (e.g. players in a soccer game) but introduce this information as soon as content is paid for.
- P2P streaming: introduce SVC as a tool to take advantage of information multiple-source distribution.
- User generated content: typical content accessed by heterogeneous terminals and through different networks, which could make the most of SVC.
- Provisioning and video preparation: how to analyse, index, pre-process, check and assign DRMs to a single video within its different layers.
- Video mail server optimization: allow access to only the lower version of the video when transmitted through a mobile network, and full resolution when viewed via a residential access, or appropriate intermediate versions according to network and terminal conditions (SVC saves storage of intermediate versions).
- Handheld terminal battery optimization: switch to lower resolution if autonomy (battery life) is lower than a predefined threshold (and inform the customer that he or she still has a given number of minutes left for visualization at the current resolution, and a bigger number of minutes at a lower resolution).
- Heterogeneous terminals and access networks for videoconferences, e-learning and video surveillance: SVC allows us not to impose the lowest network and terminal capacities on the rest of the participants.

Definitions

720p/50 High-definition progressively-scanned TV format of 1280 x 720 pixels at 50 frames per second

1080i/25 High-definition interlaced TV format of 1920 x 1080 pixels at 25 frames per second, or 50 fields (half frames) every second

1080p/50 High-definition progressively-scanned TV format of 1920 x 1080 pixels at 50 frames per second

- Efficient signal monitoring on contribution links by decoding only a low resolution instance of the signal instead of decoding the full video stream.

4. Technology evaluation

The MPEG committee defined a “requirement” document for SVC prior to the actual standardization process, in which the most important requirements were:

- To provide a standard compatible with the state-of-the-art, i.e. H264/AVC. This point has been fully addressed since each SVC file base layer is H264/AVC encoded;
- To compress, in a single stream, different versions (e.g. different resolutions or different qualities) of the same video:
  - more efficiently than if the different versions are separately H264/AVC-encoded, then used together (Simulcast) – this point is addressed;
  - with 10% maximum of additional information compared with an H264/AVC encoded stream of maximum resolution/quality. This point depends on the use case scenarios,. Basically, the more levels included, the better gain compared to separate compressions, but the more diverse the spatial layers, and extra information is required.

Assessing a new technology means defining tests to evaluate its performance against alternative existing solutions in the same context. Defining SVC tests is not that easy because it is not a new compression standard that you can compare to another older one, but a new way to compress multiple representations of the same information. Next, once the target application is chosen, we need to define the embedded representation of the information, i.e. the different ways the video can be exploited: spatial enhancements (at which resolutions?), quality enhancements (up to which quality?), temporal enhancements (which frequencies?) ... or a mix of all these enhancements?

All of this implies that the comparison depends on the targeted application:

- For multicast service environments, we can compare an SVC stream transmission to the sum of the simulcasted information encoded with H264/AVC as illustrated in *Fig. 10* on the next page. Simultaneous transmission of the yellow-coloured streams (on the left of the diagram) has to be compared with transmission of the unique blue-coloured stream on the right – keeping in mind that it requires stream adaptation/extraction to be performed on it, somewhere between the stream production area and the end device.
- For VoD services, we can compare SVC streams with the sum of the encoded stored files (storage of the yellow-coloured streams compared to storage of the light blue-coloured stream, noting that the indexing is easier in the "multiple-versions-in-a-single-file" solution).
- For content portability, we can compare the SVC capability with the transcoding applied to a first encoded version of the video.

SVC performance evaluations on particular use cases have been conducted by different research labs, industries, and the JVT to quantify SVC performances both objectively (metric-based) and subjectively (visual quality). A summary of those test results is described in the following paragraphs.

4.1. Performance evaluations

4.1.1. Visual quality assessment test by Orange Labs

Orange decided to state the criteria that were essential for current audiovisual services and to evaluate difficult situations in order to have a low anchor and recognize that only better results can be expected in the near future.

Orange chose the "ADSL constant eligibility level" criteria as a mandatory point for introducing new technologies; in other words, Orange decided to evaluate what it would mean to provide SVC-encoded TV and video at the current service bitrates. Orange then had to compare a single layer H264/AVC encoded-decoded HD file *(bottom left yellow file of Fig. 10)* with the completely encoded-decoded HD SVC file containing an embedded SD base layer *(blue file of Fig. 10)*.

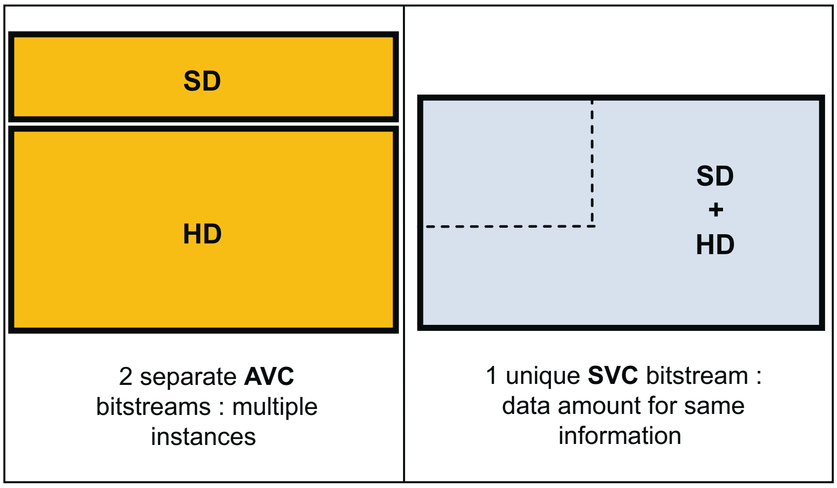


Figure 10
H264/AVC single layer vs. SVC single stream

It is important to notice that, unlike evaluating the "classical" compression method where we compare the bitrate difference for a given fixed quality, here the quality loss at constant bitrate is being evaluated. Beware, it doesn't mean that SVC decreases the quality when compared to previous methods (otherwise, what is the point of considering it to replace previous methods!). It means that this is an evaluation of the side effects caused by scalability: the extra cost required for the "maximum" version (no attention being given to the fact that "minimum" versions are then intrinsically transmitted too since they are embedded in the file) compared to the way that such a version is encoded with other methods.

Tests have been performed at different bitrates: 12 Mbit/s, 10 Mbit/s, 8 Mbit/s, 6 Mbit/s and 4 Mbit/s.

For testing bitrate *n*, an H264/AVC file was encoded completely at *n* Mbit/s, and the SVC file had a base layer for SD always encoded at 2 Mbit/s, plus an enhancement layer encoded at *n*-2 Mbit/s.

More precisely, the tested scenarios were:

a) H264/AVC (720p/50) versus SVC (576i/25, 720p/50)
b) H264/AVC (1440x1080i/25 [2]) versus SVC (576i/25, 1440x1080i/25)

These figures are summarized in *Table 2*.

Table 2
SVC/AVC bitrate figures for visual comparisons

	H264/AVC bitrate (Mbit/s)	SVC bitrate (Mbit/s)		
		Base	Enhancement	Global
1280x720p/50	4	2	2	4
	6	2	4	6
	8	2	6	8
	10	2	8	10
	12	2	10	12
1440x1080i/25	4	2	2	4
	6	2	4	6
	8	2	6	8
	10	2	8	10

2. 1440 samples per line are subsampled from a 1920 samples/line source signal.

They used MPEG-ITU JSVM (Joint Software Verification Model) to generate both the H264/AVC and SVC files.

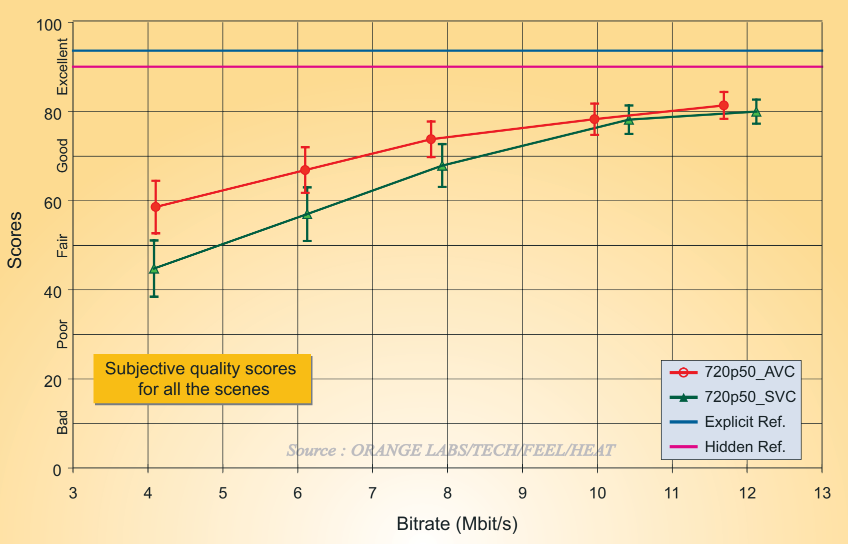
The visual quality assessment was performed on a 46-inch LCD display, following the SAMVIQ (Subjective Assessment Methodology for Video Quality) method [6][7].

Fig. 11 and *Fig. 12* show both sets of test results.

Several conclusions can be drawn from these tests:

- The SVC stream (with embedded SD and HD) provides similar to better visual quality than the single-layer HD AVC stream for bitrates greater than 7 Mbit/s. Transmitting both SD and HD as single-layer AVC would require a higher bitrate (the sum of the respective streams' bitrates) than the SVC stream.

  For an HD AVC single-layer bitrate of less than 7 Mbit/s, SVC needs, for providing a similar visual quality, an extra bitrate corresponding to the SD AVC single-layer one.



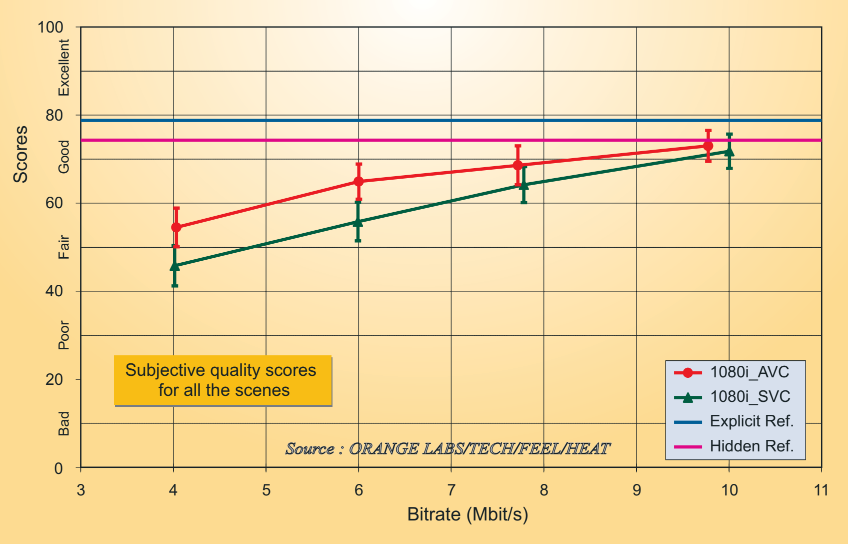


Figure 11 *(upper)*
Results for tests A

Figure 12 *(lower)*
Results for tests B

- In critical cases, the loss between H264/AVC single representation and SVC maximum representation can reach up to 10 MOS (Mean Opinion Score) points, which is a noticeable difference.
- In non-critical cases, the difference is always noticeable, but not penalizing, since the range of appreciation is generally maintained (e.g. "excellent", "good").
- SVC performs better if the base layer is of good quality since all enhancement predictions rely on it.
- If a sequence is originally interlaced, then SVC performs better with interlaced-to-interlaced scalability than when mixing interlaced with progressive layers.
- If a sequence is originally progressive, then SCV performs better with progressive-to-progressive scalability than when mixing interlaced with progressive layers.

Please note that all these results have been obtained with only the publicly-available versions of SVC software (MPEG/ITU JSVM). Needless to say, these versions are not as optimized as future industrial implementations will surely be when available.

The tests were designed with the very precise goal of identifying the impact on quality when replacing H264/AVC with SVC at constant eligibility (i.e. constant bitrate, CBR) for current services. This constant bitrate affects the highest level, here HD resolution, since we impose that the SD base layer within the SVC file is encoded at today's SD eligibility level, i.e. 2 Mbit/s. This is indeed the way to ensure compatibility with existing services because the SD layer can simply be decoded by those having an H264/AVC currently-deployed decoder and no SVC decoder. So we simulate here a base that fulfils current TV service requirements.

4.1.2. Performance evaluation by Thomson's Corporate Research

Since January 2005, H.264/AVC-based HD broadcasting has been rolling out in the market with millions of receivers being shipped (DirecTV, Echostar, BSkyB,...). The HD formats are either 720p/50-60 or 1080i/25-30.

For premium services (e.g. Sports), broadcasters want to broadcast 1080p/50-60 while maintaining the existing customer base. Broadcasting of 1080p/50-60 SVC bitstreams enhances video resolution from 720p/50-60 or 1080i/25-30. Enhanced set-top boxes could be provided to premium customers, with additional 1080p/50-60 SVC (and H.264/AVC) capability – without the need to exchange recently shipped set-top boxes.

a) 720p/1080p versus 1080i/1080p comparison

Today, broadcasters are mainly using 1080i/25-30 for HDTV. Tomorrow, source capture will be more and more 1080p/50-60. SVC enables a potential migration towards 1080p/50-60 using a backward-compatible solution with H264/AVC (base layer). The main remaining question for the H.264/AVC base layer is either to encode it as 720p/50-60 or 1080i/25-30? Thomson Corporate Research has performed a first objective assessment using the SVC reference software and looking at rate distortion curves.

Test conditions

We have used the JSVM software for a 2-layer case:

- either 720p50/1080p50: progressive-to-progressive inter-layer prediction;
- or 1080i25/1080p50: interlace-to-progressive inter-layer prediction using coded field pictures.

The encoder settings were as follows:

- Scalable High profile;
- Hierarchical GoP size 4 (I_bB_bP coding structure);
- Intra period = 32;
- 100 first frames;
- H.264/AVC base layers encoded at fixed (constant) bitrates (CBR)
  - R0 = 4 Mbit/s;
  - R1 = 6 Mbit/s;
  - R2 = 8 Mbit/s;
  - R3 = 10 Mbit/s;
- For each rate, SVC enhancement layers were encoded using quantization parameter QpEL = QpBL + {0, 4, 8, 12}.

Results

Some results of these tests are given in *Figs 13* to *18* on the next two pages.

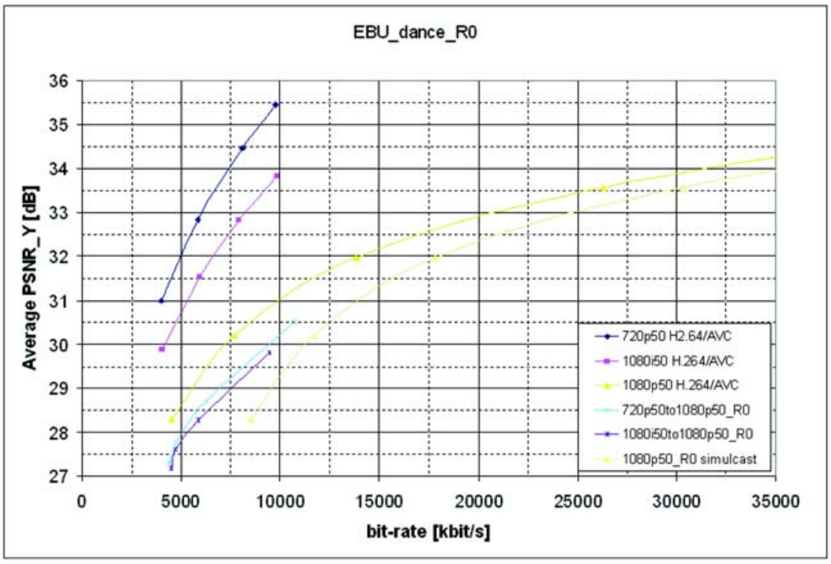


Figure 13
Results for sequence EBU_dance at rate R0

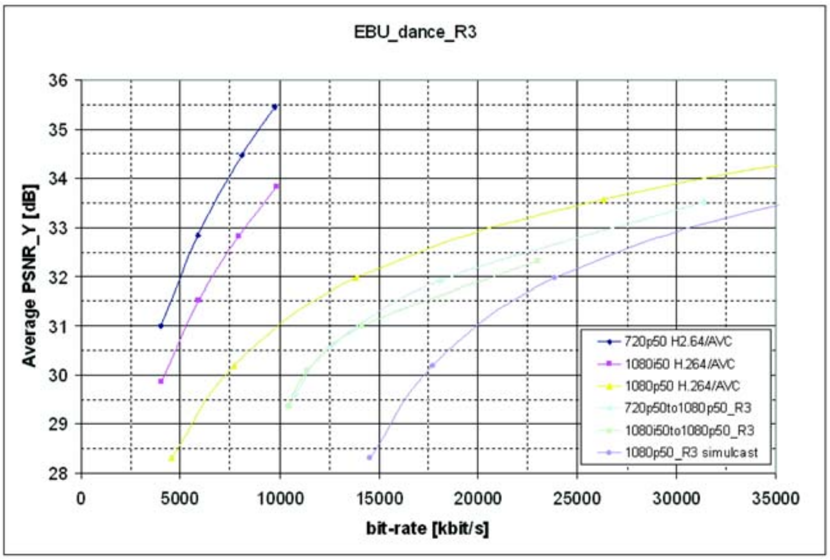


Figure 14
Results for sequence EBU_dance at rate R3

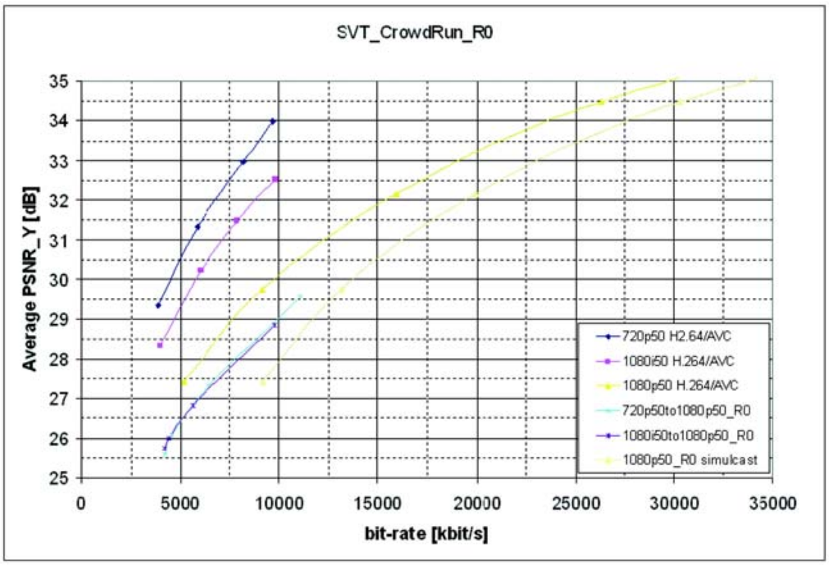


Figure 15
Results for sequence SVT_CrowdRun at rate R0

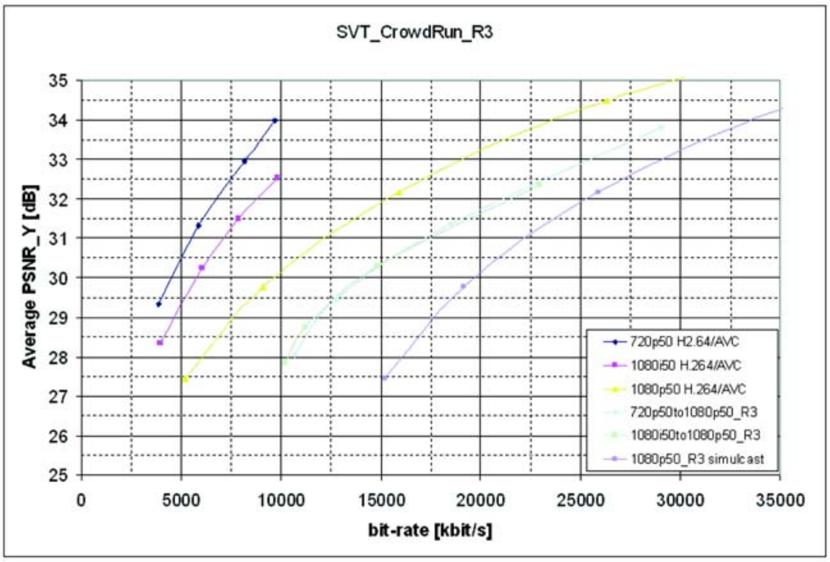


Figure 16
Results for sequence SVT_CrowdRun at rate R3

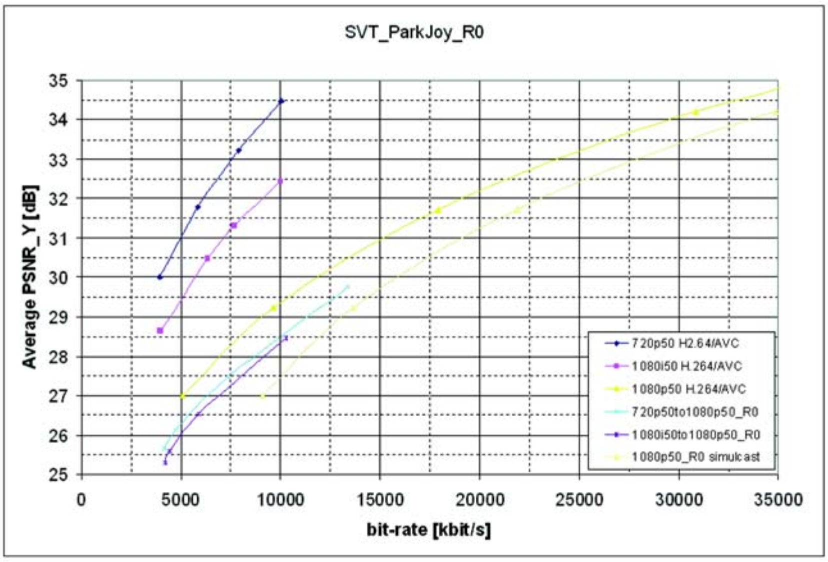


Figure 17
Results for sequence SVT_ParkJoy at rate R0

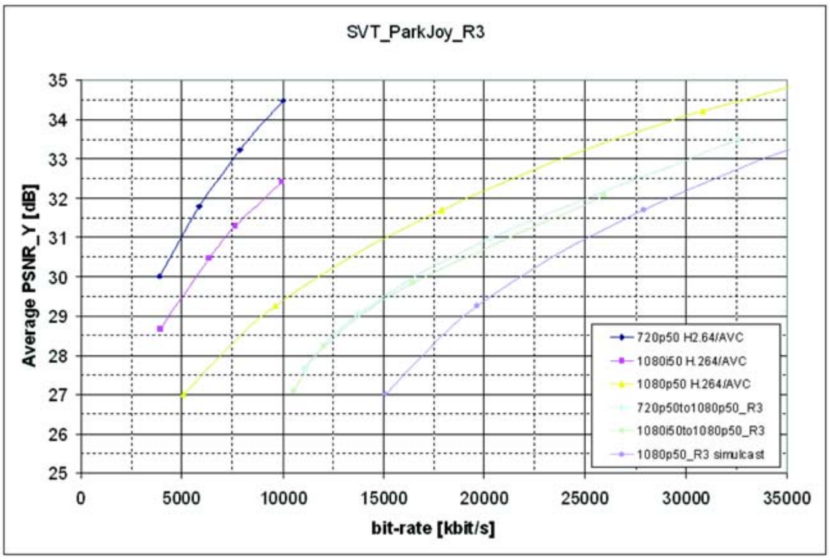


Figure 18
Results for sequence SVT ParkJoy at rate R3

Conclusions

Using an H.264/AVC 720p/50 or 1080i/25 base layer with a 1080p/50 enhancement layer provides approximately the same PSNR (Peak Signal to Noise Ratio) results, in terms of compression efficiency (a difference of less than 0.5 dB). Furthermore, it can be noted that the gain in bitrate compared to simulcast is shown and confirmed (~30%).

Knowing that PSNR doesn't always well correlate with visual perception, further subjective visual quality assessments should be conducted on a wider set of test sequences to assess the PSNR results.

4.1.3. Performance evaluation by JVT

Early this year, the JVT group released a report on SVC performance evaluation over a set of typical test cases [8]. The tests included:

- objective quality evaluation using the PSNR and SSIM (Structutral SIMilarity metric) methods
- subjective quality evaluation using two assessment methods based on ITU-R Recommendation BT.500. [9]

It has to be noted that the different test scenarios only considered progressive-to-progressive image-format scalability, while discarding interlaced-to-progressive (SDI to 720p/50 or 1080p/50), or progressive-to-interlace (less probable), or even interlaced-to-interlaced (SDi to 1080i/25) scalability, which might be valuable for the broadcasting industry.

The overall conclusion of the test was that SVC will provide a 17 to 34% bitrate gain compared to simulcast, depending on the application. On the visual quality side, for critical sequences it needs up to 10% more bitrate to achieve similar or even better quality than the single-layer stream. For further information, please refer to the following document [8].

4.2. Implementation and complexity

At the moment, only the reference software implementation, JSVM, of the standard (now in version 9.12) is freely available. Research labs such as HHI or the industry are constantly developing optimized versions of the encoders.

Encoder-complexity evaluations still need to be done, even if it doesn't seem be an issue with the software version.

Additional decoder complexity in comparison with H.264/AVC is believed to be limited. Indeed, the SVC design specifies a single motion compensation loop (by imposing constrained intra prediction within reference layers). Thus, the overhead in decoder complexity for SVC compared to single-layer coding is smaller than that for prior video-coding standards, which all require multiple motion compensation loops at the decoder side. Additionally, each quality or spatial enhancement layer NAL unit can be parsed independently of the lower layer NAL units, which may further help in reducing the decoder complexity.

In order to keep track of the changes in software development and to always provide an up-to-date version of the JSVM software, a CVS server for the JSVM software has been set up at the *Rheinisch-Westfälische Technische Hochschule* (RWTH) in Aachen, Germany. The CVS server can be accessed using WinCVS or any other CVS client. The server is configured to allow read access only, using the parameters specified in *Table 3*. Write access to the JSVM software server is restricted to the JSVM software coordinators group.

Table 3
CVS access parameters

authentication:	pserver
host address:	garcon.ient.rwth-aachen.de
path:	/cvs/jvt
user name:	jvtuser
password:	jvt.Amd.2
module name:	jsvm or jsvm_red

5. Standardization information

As the audiovisual ecosystem has evolved in multiple dimensions (services, terminals, network access) and the services may also be rendered through retail devices, the need today for interoperability has never been so critical. Standardization bodies become interested in a technology when many service operators and industrial companies begin to take an interest in it. Thus, many standardization bodies and open forums have announced working groups on SVC:

- **DVB:** DVB is of considerable influence in the world of audio/video codecs, and the DVB H264/AVC groups have included SVC in their work programmes for mobile broadcast (DVB-H), 1080p and IPTV. The IPTV groups are now considering SVC in their content-delivery task forces.
- **3GPP:** SVC will be presented for its impact on architecture.
- **Open IPTV Forum:** since the first phase relies on currently available codecs and is near to being completed, a second phase will be started soon and we have offered to include SVC in working topics for both device codecs and IPTV infrastructure.
- **ITU Focus Group IPTV:** liaisons are active with DVB on the codec topics.
- **EBU:** the Delivery Management Committee (DMC) started a project group D/SVC in June 2008 to investigate the objective and subjective performance of SVC in broadcasting. The group is open to both EBU Members and the industry.

There is still a need to see if:

- **DLNA** (Digital Living Network Alliance, home network standardization activities) would be interested in including SVC as part of a work item. We believe so, since today there is a considerable lack of standardization for retail devices. DLNA might be a good candidate since it deals with home networking matters.
- **TISPAN** would need to investigate if and how SVC may be of impact over the next generation network architecture it is defining, mostly in terms of flows between functional blocks, not only in the content traffic plane, but also in the control one.
- **ATSC** (US) and **DMB** (Korea) have announced doing some work on SVC, or have included requirements for scalability functionalities.

6. Conclusions

SVC seems to be of interest and very promising for current and future audiovisual services, especially in the area of:

- Video-on-Demand, to reduce the costs otherwise associated with the generation of multiple formats of the same video, especially for the long-tail management.

- Vector for efficient dissemination of premium HDTV services (1080p/50-60).
- Mobile video transmissions, to ensure a better continuity of service.
- Finer adaptation to home parameters such as device characteristics (HD screens) and available network access bandwidth.
- Quality of experience such as channel-switching time, fast forwards and rewinds.
- Dynamic bandwidth access management.

Adi Kouadio obtained a B.Sc. and M.Sc. in communication systems at the Swiss federal institute of Technology (EPFL) in Lausanne, Switzerland. After working for Fastcom Technology SA on several projects relating to video-based object recognition, he joined the EBU Technical Department in 2007. Here, he is heavily involved in studies and evaluations of compression systems for various broadcasting applications (HDTV production, contribution and distribution).

On behalf of the EBU, Mr Kouadio liaises with the ISO/IEC JTC1/SG26/WG1 (JPEG) and ISO/IEC JTC1/SG26/WG11 (MPEG) working groups.

Maryline Clare graduated from INSA (*Institut National des Sciences Appliquées*, an engineering school in France) in 1989. She gained 10 years of still-image coding experience at Canon Research France and was heavily involved in the JPEG2000 standardization effort, during which time she was both "Head of the French Delegation" and the "Transform - Quantization - Entropy coding" group co-chair.

Ms Clare has continued working in the advanced video compression domain, first in Canon Research France then in Orange Labs, France Telecom R&D in Rennes. Starting with the AVC standard (Advanced Video Coding, MPEG-4 - Part 10, ITU H.264), she then closely followed its scalable extension SVC (Scalable Video Coding), on which she is currently leading a project in Orange Labs.

Ludovic Noblet received an Electronics and Computing Systems Dipl.-Ing. from the *Ecole Polytechnique de l'Université de Nantes*, France in 1992. He started his career at Alcatel where he was in charge of introducing internet technologies within the Alcatel private network (Alcanet). He then moved to Thomson Corporate Research in October 1994 where he was involved in designing and contributing to the development and success of three successive generations of MPEG-2 encoders and one generation of MPEG-2 decoder. In 2002, he started working on the very first H.264/MPEG-4 AVC encoding implementations for Thomson's first generation of SDTV and HDTV AVC encoders.

In 2004, Mr Noblet joined France Telecom as an IPTV architect and senior technical advisor for the introduction of H.264/MPEG-4 AVC and high-definition within the Orange TV service. In September 2006, he was appointed head of the "Advanced Video Compression" team at Orange Labs.

Since December 2004, Mr Noblet has been the Orange representative at DVB, in both commercial and technical ad hoc groups for defining the use of A/V codecs in DVB applications. He also represents Orange at the Open IPTV Forum on the same topics.

Vincent Bottreau received the french state degree of Physics Engineer from the *Ecole Nationale Supérieure de Physique Strasbourg* (ENSPS), France, in 1999. He also received the DEA (diploma validating the first year of a Ph.D. programme) in Photonic and Image Processing from the *Université Louis Pasteur*, Strasbourg, France, in 1999.

From 1999 to 2003, Mr Bottreau was with Philips Research in Suresnes, France where he worked as a Research Scientist in the Video and Communication group. In 2003 he has joined IRISA as a Research Expert in the TEMICS team. Since 2005 he has been an R&D engineer at Thomson R&D, France. His research activities include video coding, with a special focus on motion estimation, scalability and transcoding. He is in particularly involved in the ITU and MPEG standardization activities and, more specifically, on SVC.

- Open Internet IPTV applications.

SVC is a good candidate technology for achieving the “content anytime, anywhere” target. However, further studies need to be conducted to further assess its efficiency. Among them we include:

- Other visual assessments to provide comparisons with alternative solutions to SVC (i.e. simulcasting, multiple description coding, transcoding).
- Reference test sequences and full source-quality video should be provided in all common media formats (CIF, QCIF, SDTV, HDTV, etc.).
- Visual assessments should be redone when optimized encoders are available because there is obviously room for improvement.
- Encoder complexity evaluation.

7. References

[1] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard and Ajay Luthra: **Overview of the H.264/AVC Video Coding Standard**
IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, pp. 560-576, July 2003.

[2] Heiko Schwarz, Detlev Marpe and Thomas Wiegand: **Overview of the Scalable Video Coding Extension of the H.264/AVC Standard**
IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Scalable Video Coding, Vol. 17, No. 9, pp. 1103-1120, September 2007, Invited Paper.

[3] ITU-T Recommendation H.262 and ISO/IEC 13 818-2 (MPEG-2): **Generic Coding of Moving Pictures and Associated Audio Information - Part 2: Video**
ITU-T and ISO/IEC JTC 1, 1994.

[4] ISO/IEC 14 496-2 (MPEG-4 Visual Version 1): **Coding of audio-visual objects - Part 2: Visual**
ISO/IEC, Apr. 1999.

[5] Edouard Francois, Jérome Viéron and Vincent Bottreau: **Interlaced coding in SVC**
IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 9, pp 1136-1148 September 2007.

[6] EBU BPN 056: **SAMVIQ - Subjective Assessment Methodology for Video Quality**
Report by EBU Project Group B/VIM (Video In Multimedia), May 2003.

[7] F. Kozamernik, V. Steinman, P. Sunna and E. Wyckens: **SAMVIQ - A New EBU Methodology for Video Quality Evaluations in Multimedia**
IBC 2004, Amsterdam, pp. 191 - 202.

[8] N9577: **SVC Verification Test Report**
ISO/IEC JTC 1/SC 29/WG 11, Antalya, TR - January 2007.

[9] ITU-R Recommendation BT.500-11: **Methodology for the Subjective Assessment of the Quality of Television Pictures**
Question ITU-R 211/11, Geneva, 2004.

[10] TD 531 R1 (PLEN/16), Draft new ITU-T Rec. H.264 (11/2007) | ISO/IEC 14496-10 (2008) Corrigendum 1: **Advanced video coding for generic audiovisual services: Corrections and clarifications**
Geneva, 22 April - 2 May 2008

Tag-based information retrieval for educational videos

Mark Melenhorst, Marjan Grootveld and Mettina Veenstra
Telematica Instituut

In this article we report on a study that explores the contribution of social tags, professional metadata and automatically generated metadata to the retrieval process. In the tagging phase of the study, 194 participants tagged a total of 115 educational videos. In the search phase, 140 participants searched the video collection for answers to eight questions.

The results show that, in the current context, social tags yield an effective retrieval process, whereas automatically-generated metadata do not. In this study we have found some evidence for the claim that social tagging is effective, because in the retrieval process the same terminology is used as in the process of assigning metadata.

1. Introduction

Due to a steadily growing amount of media content – including collections of movies, music, archived TV programmes as well as websites' content – metadata that enable users to find what they need become more and more urgent. The problem of annotation, i.e., adding metadata to content, increases likewise. Firstly, because an increasing amount of resources needs to be spent on annotation, as argued by Mathes [1]. Secondly, the quality of the metadata becomes more and more critical: a growing collection requires good indexing of the content to enable effective content retrieval. The question then is to which extent existing forms of metadata suffice to guarantee an easy retrieval of information on the one hand and to keep the costs of adding metadata to resources under control on the other hand.

Technical solutions may help to provide better retrieval support to users, while these solutions simultaneously alleviate the pressure on resources required for adding metadata. For the same reason, the notion of involving the target group is appealing, even more so since many consumers have become content producers, who are well aware of the impact of proper metadata on retrieval.

Most tagging research focuses on the act of tagging and on emerging folksonomies and communities, whereas the contribution of social tags to information retrieval is rather underexposed (e.g. [2] is an exception). We assume that tagging is an efficient method for annotating content in a way that fits a user's needs and conforms to a user's frame of reference during retrieval. In this article we report on a study of the respective contributions of ***social tagging***, ***automatic keyword extraction techniques*** and ***professional annotation*** to the retrieval process.

2. Tags as a metadata source

2.1. Introduction

Many researchers state that the common formal way of professionally assigning metadata is no longer the optimal way of annotating content, both in terms of efficiency and in terms of user support[1]. Macgregor & McCulloch [3] argue that if applied to digital libraries and the web, traditional metadata creation and indexing suffer from scalability problems and the need for a substantial amount of resources. In an opinionated web article [5], Shirky argues that “Users have a terrifically hard time guessing how something they want will have been categorized in advance, unless they have been educated about those categories in advance as well, and the bigger the user base, the more work that user education is.” In a similar vein, [6] discusses tagging as a potential means to bridge the semantic gap between art historians and art museums’ visitors.

Social tagging or ethnographic classification is an *informal* way of assigning *user-defined* labels, called “tags”, to content items. Instead of formally classifying content according to the principles of documentation experts, users define the terms themselves informally, based only on the associations that the item to be classified triggers. Definitions in this domain abound, as shown in *Table 1*.

Table 1
Definitions of social tagging

Study	Object of definition	Definition
[6]	Social tagging	“The collective assignment of keywords to resources”
[7]	Tagging	“Labelling objects with free-style descriptors”
[8]	Tag	“One-word descriptor[s] or term[s] to describe the image or bookmark”
[9]	Social tagging	“The collaborative activity of marking shared online content with keywords, or tags, as a way to organize content for future navigation, filtering or search”
[3]	Collaborative tagging	“A practice whereby users assign uncontrolled keywords to information resources”
[10]	Tags	“Short free form labels used to describe items in a domain”

One observes a variation in what is defined (i.e. the tags or the activity of tagging) and in the extent to which social aspects of tagging are included. The common denominator in these definitions is that users can create a set of descriptors to characterize content items. Furthermore, [6] and [9] add a collaborative aspect to the tagging process.

2.2. Advantages and disadvantages of tagging

The problematic scalability, the costs, and the lack of compliance with the user’s frame of reference have been introduced as the disadvantages of formal classification systems. Also, such systems are hard to keep up to date when, for instance, geographical information becomes outdated (e.g. about the German Democratic Republic) or new concepts emerge [5]. Furthermore, strictly hierarchical systems in which concepts cannot have more than one broader term do not easily allow for the combination of concepts from different parts of the classification system to express compound topics, e.g. “African butterflies”.

Since tagging is assumed to be an alternative to formal classification, its advantages and disadvantages are presented in the next section.

The hypothesized benefits of tagging are:

- tags allow for more flexibility in organizing content;

1. Note that this typically relates to descriptive metadata rather than to structural or administrative metadata [4].

- tags facilitate the annotation process because little or no knowledge about the system is required;
- tags facilitate content retrieval and discovery.

A folksonomy is the result of social tags being added to content over time. It reflects an emerged consensus about the terminology people use to describe content items. Trant ([6], p. 83) defines a folksonomy as "the assemblage of concepts expressed in such a cooperatively developed system of classification" (in contrast to [1], who prefers the notion of "categorization" to the stricter notion "classification").

Folksonomies allow for far more flexibility than strict ontological systems can provide [5]. This is beneficial for, in particular, non-expert users – both if they want to organize their content under various headings and when they intend to retrieve content. For instance, tags support the goal-driven search for a specific item.

Apart from these advantages of tagging, research also identifies a number of disadvantages to social tagging. According to Golder & Huberman [11], Mathes [1], and Guy & Tonkin [12], there are also disadvantages to tagging that require serious consideration:

- **Ambiguity** (polysemy and homonymy) – a term or concept having multiple meanings.
- **Synonymy** – multiple terms which describe the same things or actions. This includes misspellings, spelling variations, conjugated verbs and mixed uses of singular and plural forms.
- **Level of term specificity** (hyponymy and hyperonomy) – e.g. "Siamese", "cat" or "animal".
- **Tag syntax** – multiple words, spaces, symbols and other idiosyncrasies.

Solutions to these problems fall outside the scope of this article, but these are the directions that we expect them to come from:

- Combining formal ontologies with free-style tags. Whereas in strict ontological systems the user is forced to think in terms of the ontology, in tagging systems the ontology is instrumental to the free-style tags that users add to content.
- Power of the masses. Research on **http://del.icio.us** [11] demonstrates that the number of new tags attached to particular types of items declines over time, indicating that with a critical mass of users, a more or less stable system emerges.
- Using ontologies and natural language analysis. Furthermore, ontologies and natural language analysis techniques together can be used to deal with the issue of synonymy (see [13]), which relates words to hyponyms and hyperonyms, thereby reducing the term-specificity problem.

In the next section we describe a video-tagging application that takes these solutions into account. The main focus of this application however is to support content retrieval by means of social tags. The application will be used to test whether social tags fulfil the expectations with respect to their support for content retrieval. This study will be presented in *Sections 4* and *5*.

3. The video-tagging application

3.1. General description

For the purpose of this study an application called ViTa (short for video tagging) was developed that allows users to retrieve, play and tag videos. The application provides access to a limited collection of short video clips that are supplied with professional metadata such as keywords, the title, and the description of the clip. The metadata were provided by Teleblik, a Dutch organization that provides access to multimedia sources for students in secondary education. The metadata were written by professionals, who take the age and knowledge of the users into account.

Searching / browsing on the one hand and playback and tagging on the other are the two key components of the system. In this section we sketch the generic application; aspects that specifically relate to the experiment are presented in the next section.

Searching and Browsing

Users start their search by entering a search term. Each entry in the result list consists of the video clip's thumbnail, its title, duration, keywords, first lines of the description, and its tags. By clicking on a tag, that tag is added to the search filter. Similarly, users can select content terms from the title and the description and add these to the filter. *Fig. 1* provides an overview of the search interface.

Figure 1
The search interface

Tagging

When a user clicks on a video clip's thumbnail in the result set, the application presents the video playback and tagging component. We introduced tag suggestions to help users come up with relevant keywords. Other tagging applications start offering the users suggestions for potentially interesting tags. For instance, when users start typing a tag, Del.icio.us offers them suggestions for tags they have assigned themselves by means of a small list from which the user can choose a few tags. These suggestions are based on both popular tags and the tags the user has assigned himself. The ViTa application, however, computes suggestions a priori. They appear at the same time as the video clip is loaded, after being selected from the result set. The suggestions are based on various recommendation techniques:

1) most popular tags assigned to that video;
2) tags assigned to similar videos;
3) tags that the user used most frequently;
4) professionally assigned keywords;
5) keywords extracted from the professional metadata.

No information about these sources or the reliability of tag suggestions is provided to ViTa users. The reader is referred to [14] for a discussion on these tag suggestion techniques.

Fig. 2 provides an overview of the tag interface.

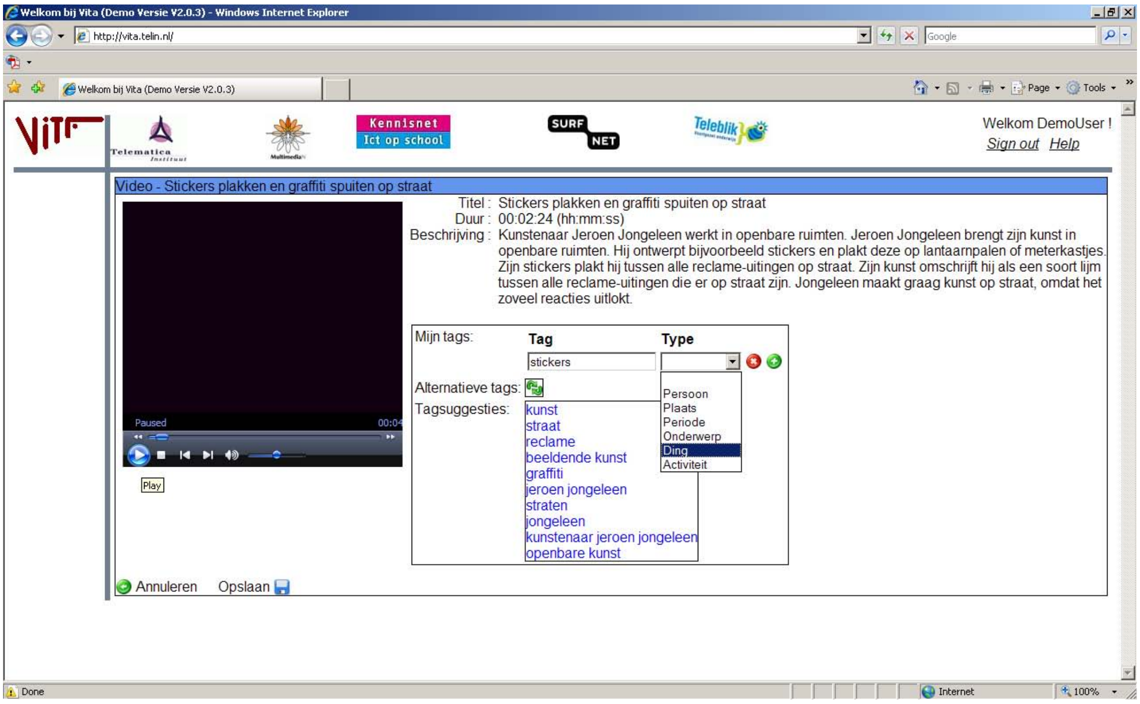


Figure 2
The tag interface

4. The search study

4.1. Research questions

Melenhorst & Van Setten [15] studied the production aspects of tagging and found large differences in the number and nature of the tags that participants assigned to TV programmes. They found that users attach meaningful tags that describe the programme's topic, express an opinion, or contain self-references. This seems to strengthen our assumption that tagging is an efficient method for annotating content in a way that fits a user's frame of reference during retrieval. In this study we verify this assumption by comparing different sources of metadata in terms of their contribution to the retrieval process. The key research question is:

> *To which extent can keywords as metadatum for educational video clips be generated by the target group, by professionals and by automatic extraction?*

In the current study, we considered successful metadata as these metadata that offer the best support for video retrieval. Whereas Van Setten & Wartena [14] discuss the part of the study concerning the automatically-derived tag suggestions, we focused on the use of various metadata in the retrieval situation.

4.2. Domain

The study focused on the educational domain, in particular on tagging and retrieving arts videos in the last classes of Dutch high schools. Since high school students are the primary target group of these videos, we assume that their tags fit the user's frame of reference during content retrieval. Therefore they have been recruited as participants [2].

2. Many so-called broadband schools from the **SURFnet/Kennisnet Innovation Programme** assisted in recruiting participants.

4.3. Terminology

The following table gives definitions of the terminology we used during the study.

Term	What it refers to
Metadata	The most generic term
Professional metadata	Professionally assigned **title** and **description** together
Keyword	Professionally assigned **keyword**
Smart keywords	Professionally assigned keywords and terms automatically extracted from all three professional metadata fields (**title**, **description** and **keyword**)
Smart tags	**Tags** that are suggested in the tagging phase
(User) tags	**Tags** assigned by users
User terms	**Terms** typed in while searching

4.4. Methodology

The study consisted of two separate phases: a tag phase and a search phase. In this article we only report on the search phase of the experiment. However, for the sake of clarity, we also outline the set-up of the tag phase.

In the tag phase, 194 participants were asked to tag 115 short video clips. Apart from the video clip itself, they could use other users' tags, professional metadata or a combination of tags and metadata. In the *BasicTagger* condition, participants could only use the metadata offered on-screen. In the SocialTagger condition, participants could make use of tag suggestions, consisting of tags that were assigned by other users. In the LazyTagger condition, the participants could use suggestions that were derived from both other users' tags and the professional metadata.

The tags that were generated were used for the search phase. The method we used for this phase of the study is described hereafter.

Tasks

In the search phase, each participant was asked to answer eight questions, the answers to which could be found in particular video clips. All participants had to answer the same questions [3], although posed in random order. The video screen contained a "this is the right video" button; when clicked, it invoked a dialogue window in which the participant could type the answer before proceeding to the next question *(see Fig. 3 on next page)*. It was possible to skip a question.

Materials

In addition to the professionally-applied metadata (title, duration, description and keywords of the clip), participants could search the collection of 115 short video clips that had been tagged previously. By manually entering a term in the search filter, participants retrieved their first results list [4]. For each result, the application shows title and duration of the clip. Depending on a participant's search condition, fewer or more of the other textual sources were available for search, which was reflected in the user interface.

3. Due to the limited size of the collection, the assignments' phrasing had to be slightly abstruse to force the participants to really search and not simply spot the correct clip.
4. The possibility to simply browse through the whole content collection was deliberately excluded.

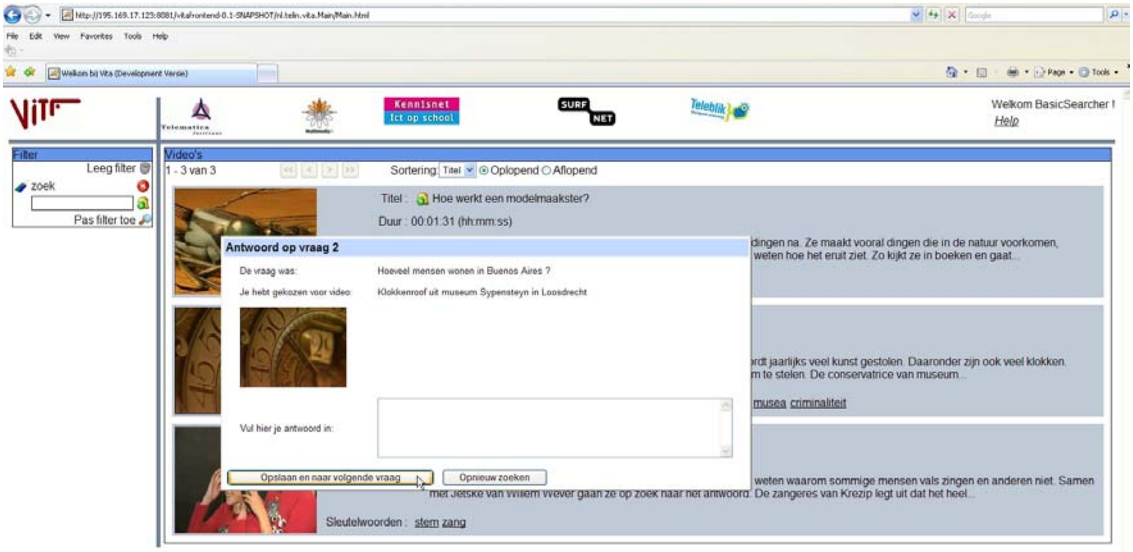


Figure 3
The answer dialogue

Participants

For the search phase we recruited 153 high school students. They participated individually, although often class-wise. No student participated in both phases of the experiment.

Experimental set-up

The participants could search various metadata sources for the clips that answered the questions posed to them. The 153 participants were randomly distributed over five groups:

- ***BasicSearcher*** could search in the professional metadata;
- ***SocialSearcher*** could search in user tags;
- ***BasicSocialSearcher*** could search in both professional metadata and user tags;
- ***DocumentSearcher*** could search in smart keywords [5];
- ***SmartSearcher*** could search in smart tags.

Table 3 summarizes the search conditions:

Table 3
Search conditions: who searches what?

Conditions	Professional Metadata	User tags	Smart keywords	Smart tags
Basic	Yes	No	No	No
Social	No	Yes	No	No
BasicSocial	Yes	Yes	No	No
Document	No	No	Yes	No
Smart	No	No	No	Yes

5. The number of available metadata in the respective search conditions varied. In order to provide DocumentSearchers with a reasonable number of metadata, the original smart keywords have been extended by means of semantic expansion ([13]).

When a participant added a term to the search filter or removed one from it and when he or she started a video, the ViTa application logged this as navigation actions. The number of navigation actions is used as an indicative measure of search efficiency.

Fig. 4 and *Figs 5* to *8* (on the following pages) show what the search results looked like in each of the conditions. In the subsequent sections we present the results of this study.

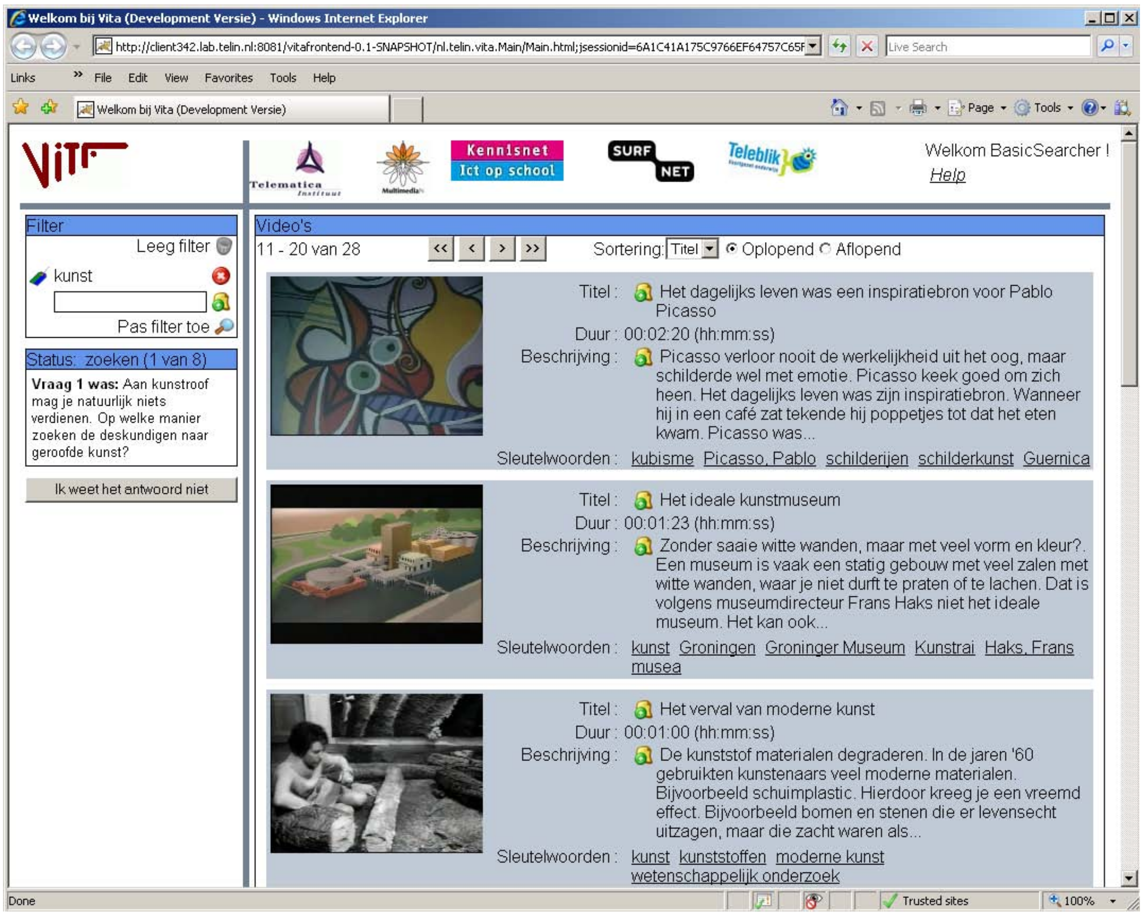


Figure 4
Search Results in the Basic Condition

4.5. Results

4.5.1. Summary of the Tag Phase

In the tag phase, we found that users produced 4373 unique tags in total for the 115 video clips. Participants were able to produce tags even when no support was offered with tag suggestions. Participants did however make use of tag suggestions, when they were offered. Those participants who could use tag suggestions did not rely only on the suggestions. They also came up with new tags themselves. Thus, the tag set that formed the input for the search phase was a mixture of tags invented by the users along with accepted tag suggestions, which in turn were derived from the metadata and from other users' tags.

In the following sections we present the results of the search phase.

4.5.2. Respondents

In total, 153 participants participated in the search phase. Unfortunately, 13 participants had to be excluded because they proved to invest too little effort in the experiment. They executed less than the minimal two navigation actions per task (entering a search term and selecting a video). As a result, they could not have given an answer and were therefore excluded.

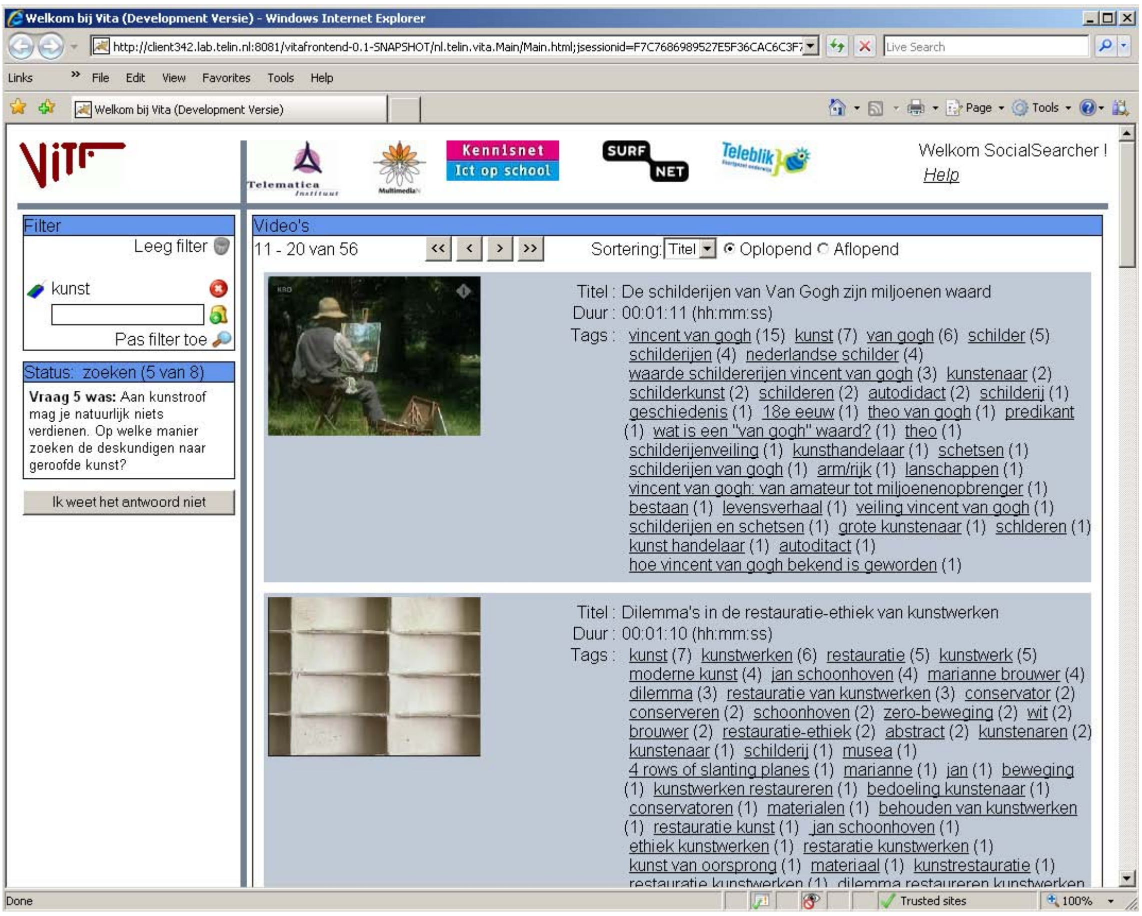


Figure 5
Search Results in the Social Condition

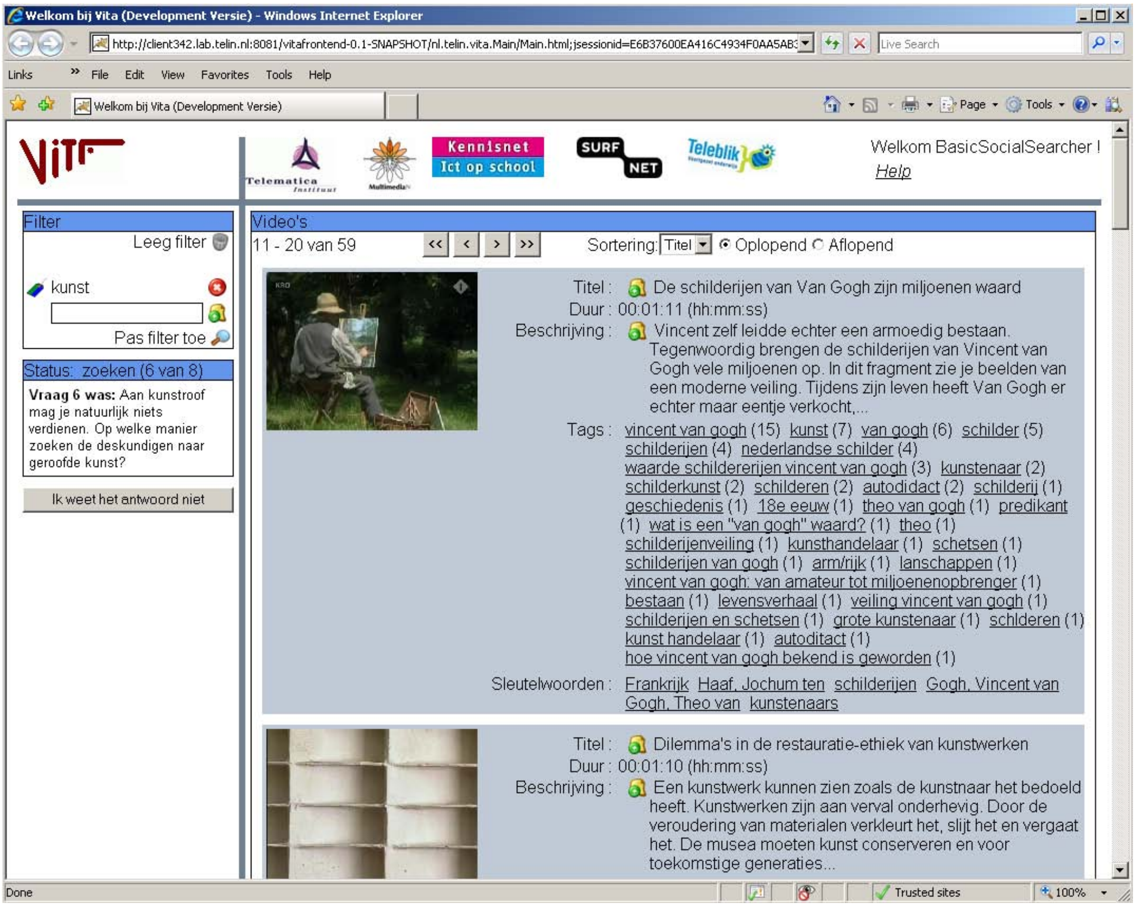


Figure 6
Search Results in the BasicSocial Condition

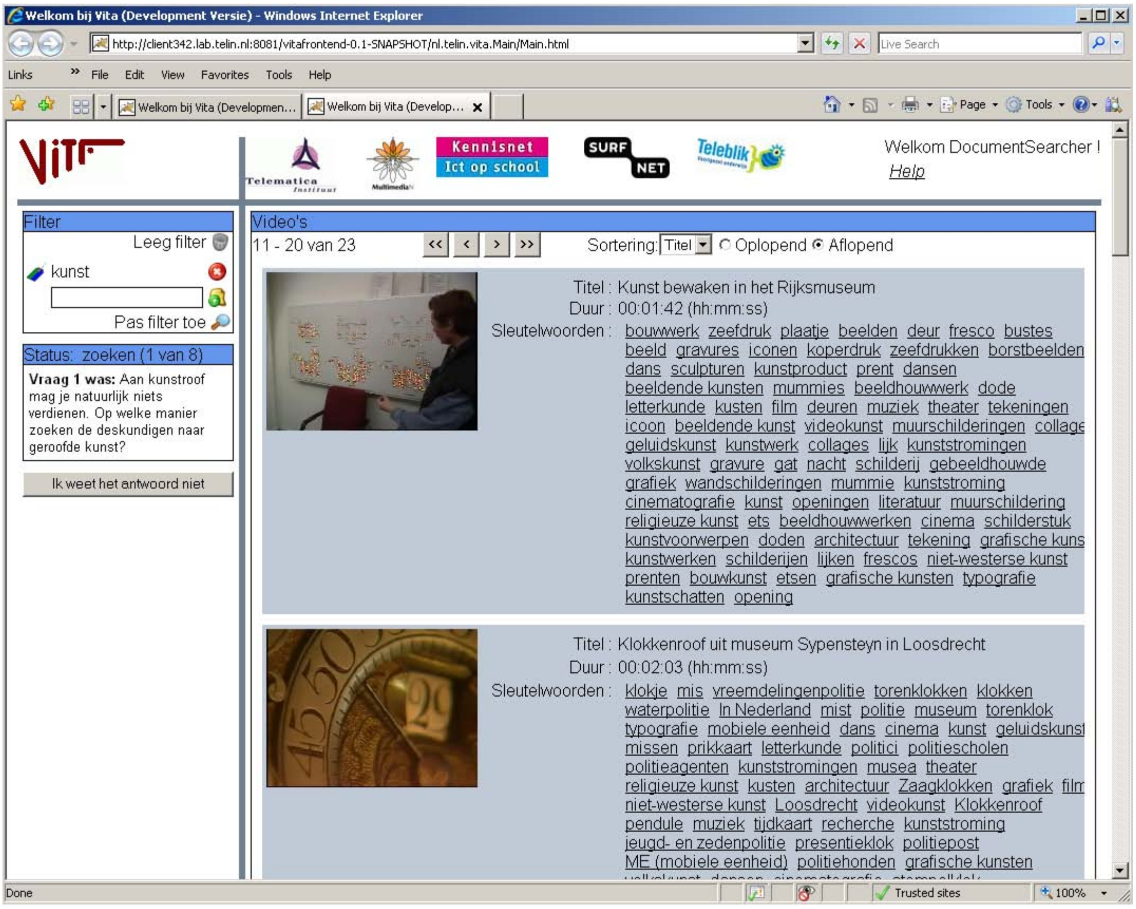


Figure 7
Search Results in the Document Condition

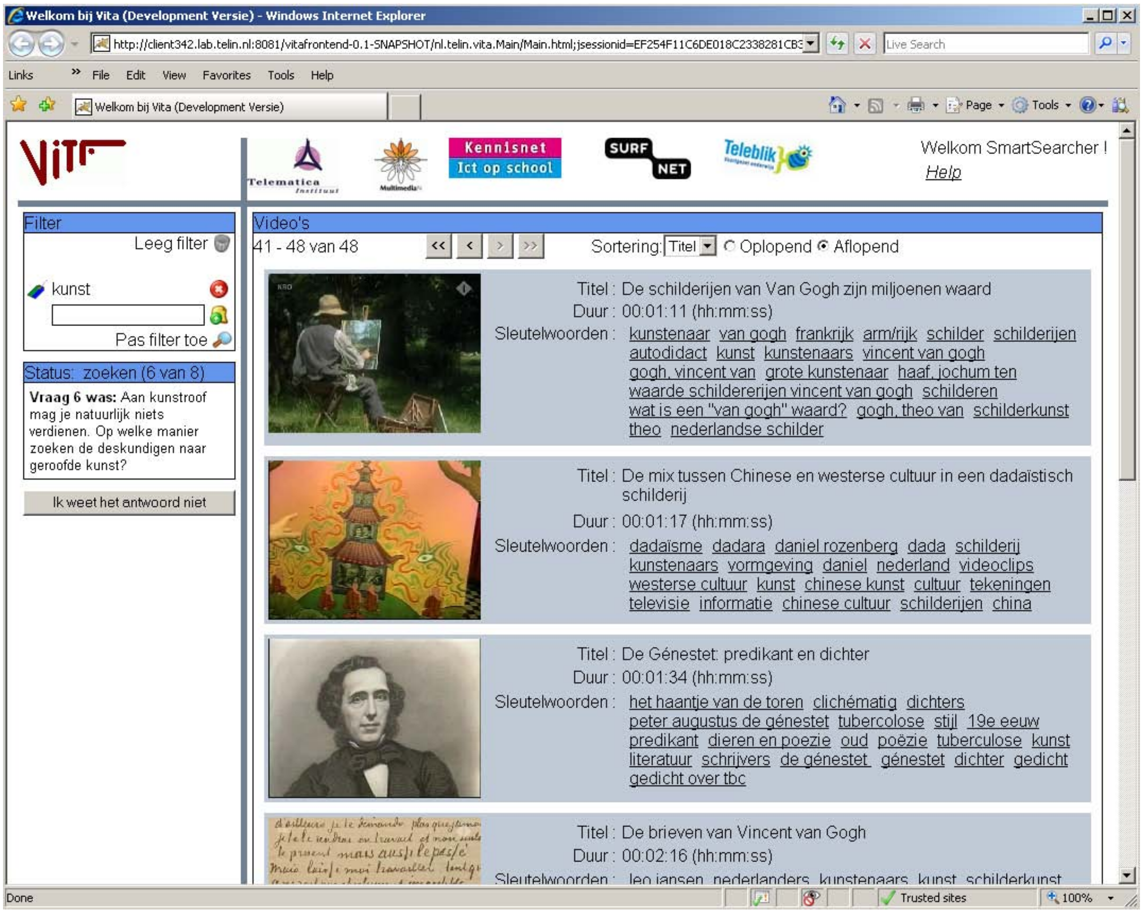


Figure 8
Search Results in the Smart Condition

Table 8 shows the distribution of the remaining 140 participants across the five search conditions.

Table 8
Number of Participants in the Search Phase

Condition	No of participants
BasicSearcher	25
SocialSearcher	31
BasicSocialSearcher	22
Document Searcher	31
SmartSearcher	31
Total	140

In *Table 9* we summarize the characteristics of the participants.

Table 9
Characteristics of Participants in the Search Phase

Background Characteristic	Mean, #	S.d.
Age	15.6	0.8
Sex		
Male	68	
Female	72	
Education (all high school)		
Havo (intermediate level)	79	
Vwo (higher level)	61	
Experience with tagging and video applications [a]		
Flickr	1.1	0.6
Del.icio.us	1.1	0.4
YouTube	4.3	1.3
Amazon	1.3	0.7
FabChannel	1.2	0.6
GoogleVideo	2.6	1.4

a. Scale ranges from 1 (never) to 6 (very often)

We did not find any statistically significant difference between the conditions on each of these characteristics ($F(4) < 2.02$; $p > 0.05$).

4.5.3. Responses to the Search Tasks

In the introduction we conceived "successful" metadata as the metadata that offer the best support to find the required information. To test whether manipulating the type of metadata affects search performance, we computed the average number of answers and the average number of correct answers for each of the search conditions. In *Table 10*, the number of questions answered and the number of correct answers is displayed, categorized by role.

Table 10
Average number of answers and correct answers

Role	*N*	Answers found		Correct answers	
		Mean	*S.d.*	*Mean*	*S.d.*
Basic *	**25**	6.4	2.3	4.5	2.2
Social * +	**31**	6.7	1.4	5.5	1.6
BasicSocial * +	**22**	6.7	1.6	5.0	1.6
Document	**31**	5.3	2.2	3.5	2.1
Smart +	**31**	6.2	2.2	4.8	2.0
Total	**140**	**6.2**	**2.0**	**4.6**	**2.0**

* Significant difference with DocumentSearcher condition for no. of answers found
\+ Significant difference with DocumentSearcher condition for no. of correct answers

As can be seen from *Table 10*, both the number of answers and the number of correct answers is lower in the DocumentSearcher condition than in other conditions, while the SocialSearcher and the BasicSocialSearcher seemed to give the most answers and the best answers.

We tested whether the participants' search condition affected the number of answers given and the number of correct answers. Both ANOVA tests yielded statistically significant differences ($F(4) = 2.53$, $p < 0.05$ and $F(4) = 4.26$; $p < 0.01$) respectively). The Tukey LSD post-hoc tests confirmed that participants in the DocumentSearcher condition answered less questions than participants in the BasicSearcher, SocialSearcher and BasicSocialSearcher conditions (Tukey LSD mean diff. < -1.10, $p < 0.05$).

In the case of correct answers, we found statistically significant differences between the DocumentSearcher condition on the one hand and the SocialSearcher, BasicSocialSearcher and SmartSearcher on the other (Tukey LSD mean diff. < -1.26, $p < 0.05$). Between the other conditions, no statistically significant differences were found.

These results suggest that in the DocumentSearcher condition, the available metadata (i.e. the smart keywords) did not correspond with the terms that users had entered to search for the answers. In other words, there seems to be a terminology gap: users seem to employ a different set of terms than professionals do. If participants search in their own terms and the searchable metadata does not contain these terms, the participants will not get the desired results.

Furthermore, *Table 10* reveals that participants in the SocialSearcher condition answer as much – or even more – questions correctly than participants who can make use of professional metadata. This provides support for the claim that social tagging can substitute or at least complement professional annotating (as explained in the introduction).

Next, we analyzed the number of navigation actions the participants needed to find the correct answer for those cases in which they succeeded in finding the correct answer. The results are shown in *Table 11*.

As can be seen from *Table 11*, the number of navigation actions is comparable to the number of correct answers. That is, in those conditions in which relatively few correct answers were given, people also needed more navigation actions.

Table 11
Average number of navigation actions to the correct answer

Condition	N [a]	Mean	S.d.
BasicSearcher	141	8.9	10.3
SocialSearcher [b]	193	6.9	7.8
BasicSocialSearcher [b]	158	6.8	7.3
DocumentSearcher	136	10.8	14.5
SmartSearcher [b]	186	7.5	8.0
Total	**814**	**8.0**	**9.7**

a. Cases represent user-search task combinations

b. Significant difference with DocumentSearcher condition

Furthermore, similar statistically-significant differences between the conditions were found with respect to the navigation actions ($F(4) = 4.71$, $p < 0.01$). The DocumentSearcher condition is again outperformed by the other conditions except for the BasicSearcher. The lower number of navigation actions for the SocialSearcher and BasicSocialSearcher proved not to be statistically significant from the other conditions.

These results suggest that users need less navigation actions when the metadata they can search comes from other users: the conditions in which tags are available for searching outperform the conditions in which professionals have generated the metadata.

4.5.4. Source of search terms

In the previous section we have analyzed the number of answers on the search tasks. In this analysis, we have not yet assessed the participants' search strategies to the extent that these strategies can be observed from the metadata elements used in their queries.

In this section we analyze the search terms the participants used to compose their queries. The 140 participants in the search phase searched with 3744 search terms in total. These search terms came from the search question, the professional metadata, the tags, the smart keywords or the smart tags, depending on the condition and hence on the type of information that could be searched. The types of information that could be searched are referred to as "sources".

In *Table 12* the distribution of the search terms is shown. As terms can occur in more than one source, we constructed a cross-table to display the source of the search terms. The cells thus represent search terms that occur in one or more sources. For instance, 202 search terms occur in both the professional metadata and the search questions simultaneously.

As can be seen from *Table 12*, 2620 of the 3744 search terms were related to either the search question or the sources of information. The remaining 1124 terms were invented by the users themselves.

Furthermore, a significant number of the search terms proved to be derived from the search questions. Apparently, the participants' search strategy primarily involved selecting the most characteristic terms from the search questions. In the current context, the effectiveness of this strategy is doubtful since the questions were formulated with the purpose of having participants really search for an answer.

In sum, 621 search terms occurred in the tags, a significant share of which also occurred in the search questions (281). The other sources cover, at the most, 379 of the 3744 search terms.

Table 12
Distribution of search terms over sources

Source	Search question	Prof. metadata	Metadata + search question [a]	Tags	Smart keywords	Smart tags	Total
Search question	941						926
Prof. metadata	187	132					351
Tags	302	54	119	146			621
Smart keywords	221				158		379
Smart tags	243					100	343
Total	1894	186	119	146	158	100	2603
Undetermined							*1141*
Total no. of search terms							**3744**

a. This column represents search terms that occur in three sources simultaneously, namely the metadata, the search questions and the tags

Since the amount of searchable information differs between conditions, it may be that the percentage of search terms that do not occur in the information sources also differs between conditions. We expected the DocumentSearchers and the SmartSearchers to have a larger percentage of such search terms. In *Table 13* we show these percentages across all conditions. We excluded the terms from the task description since these terms are not searchable sources of information that would yield search results. In the fourth column we display the total number of terms that could be searched in a particular condition.

Table 13
Average percentage of undetermined search terms

	Mean	S.d.	Search base
Basic	53.8	14.8	1932
Social	48.9	14.7	4373
BasicSocial	41.2	17.1	5466
Document	56.6	12.5	4022
Smart	54.1	13.0	1528
Total	**51.5**	**15.0**	

Note: Search base refers to the unique number of terms that could be searched

As can be seen from *Table 13*, the percentage of search terms that did not occur in the metadata sources was the lowest for the BasicSocialSearcher and the SocialSearcher, the conditions in which tags could be searched. The conditions in which metadata or extracted metadata could be searched yielded a higher percentage of terms that were not covered by any of the metadata sources. A one-way ANOVA analysis shows that the differences between the conditions are statistically significant ($F(df=4) = 4.23$; $p < 0.01$).

This suggests that user tags provide a better coverage of the search terms our participants have used than the other conditions, in particular the conditions in which only (extracted) metadata is provided (BasicSearcher and DocumentSearcher). Metadata alone – whether automatically extracted or not – seems to be insufficient to cover the terms the participants wanted to use to search for video clips. However, we have to keep in mind that the number of terms that could be searched in the BasicSearcher condition was relatively low. Thus, with social tags, users have a higher chance of successful retrieval, which is in fact an advantage of social tagging.

5. Conclusions

The results described in the previous section draw attention to a number of issues concerning the effect of different metadata sources on the retrieval process. We conclude from our analyses that, in the current context, social tags support users in finding the information they need. The results show that social tags result in an equally effective if not more effective search process than professionally-generated or automatically-generated metadata.

In the DocumentSearcher condition, in which automatically-generated metadata could be used, the smallest number of correct answers was found, while the percentage of search terms that did not occur in the metadata source that the participants could search, was highest. To a lesser extent, these results also apply to the BasicSearcher, in which only professional metadata could be used.

Dr Mark Melenhorst is a researcher at the Telematica Instituut. He holds a Ph.D. in the field of communication science (focusing on new media and communication technology). His Ph.D. thesis describes a field study on the role of note-taking software. His expertise involves technical and professional communication, interface and interaction design, user-centered design, and experimental research.

Mark Melenhorst's current work involves empirical research regarding the relation between social tagging and recommender systems, the contribution of social tagging to information retrieval processes, and users' motivation to tag video content. Apart from his current work, his research interests include user-centered design methods and applications, e-government and Web 2.0.

Dr Marjan Grootveld obtained her Ph.D. in computational linguistics from Leiden University in the Netherlands. She has developed computer-based training programmes for major German companies and is currently a member of the scientific staff of the Telematica Instituut in Enschede (the Netherlands). Her research interests centre around content engineering, learning and knowledge management. Application areas include education, cultural heritage, healthcare and industry. She has managed projects ranging from knowledge retention in the defence industry to sharing patient information in complex healthcare chains.

In Marjan Grootveld's experience, demand-driven user-oriented research is usually very gratifying. She is convinced, however, that the impact and quality of innovation projects would increase – and become even more gratifying – if more and earlier attention would be paid to the intended end users and to embedding the innovation in existing workflows.

Dr Mettina Veenstra is a senior member of the scientific staff at the Telematica Instituut in Enschede, the Netherlands. She holds a Ph.D. in computational linguistics from the University of Groningen. After finishing her Ph.D. she worked at a small multimedia publishing company as a researcher and project manager.

At the moment Mettina Veenstra is head of the Media Interaction group at the Telematica Instituut, a group of ten researchers and three software engineers. This group collaborates with partners from industry and academia on solutions for improving the access to and interaction with large quantities of multimedia content, especially in the (new) media, medical and cultural heritage domains. The focus of the research is on recommender systems, user interaction, social tagging and automatic extraction of metadata. She is involved in several national and European projects in these areas.

The explanation for the lower performance in these conditions seems to be that users do not search in terms of the automatically-generated keywords, the only source that could be searched in this condition. Consequently, within the current context, professional metadata or automatically-extracted and semantically-broadened keywords alone do not seem to be effective in supporting the retrieval process yet.

In contrast, the SocialSearcher and the BasicSocialSearcher answer the highest number of questions correctly, while the percentage of terms that did not occur in the metadata sources available for searching was the lowest. In both conditions, tags could be used. In contrast to the DocumentSearcher condition, in these conditions, participants search with terms that occur in the available metadata sources more often: the search terms corresponded with the tags.

In fact, this correspondence in terminology is a basic claim of social tagging in general: social tagging enables annotating content in the terminology of the user, which makes the search and retrieval process more effective.

As a final remark, we have to keep in mind that these results were achieved in a context where professionals have explicitly attempted to take the language used by their target group into account. In most situations this will not be the case, which may result in larger differences in language use between the metadata and the users.

6. Acknowledgements

This study was carried out within the MultimediaN [6] PERSIS project (BSIK 03031), which is sponsored by the Dutch government. PERSIS participants are the Telematica Instituut, Kennisnet Ict op school, SURFnet, Eindhoven University of Technology, FabChannel and Roessingh Research and Development. In particular the authors thank Karin van Bakel (Kennisnet Ict op school), Teleblik and the broadband schools for their cooperation.

7. References

[1] 9-A. Mathes: **Folksonomies – cooperative classification and communication through shared metadata**
Computer Mediated Communication, LIS590CMC (Doctoral Seminar), Graduate School of Library and Information Science, University of Illinois Urbana-Champaign.

[2] 3-A. Davis: **Why tagging is expensive**
Retrieved from: **http://blogs.talis.com/panlibus/archives/2005/09/why_tagging_is.php**.

[3] 7-G. Macgregor and E. McCulloch: **Collaborative tagging as a knowledge organization and resource discovery tool**
In Library Review 55 (5), pp. 91-300.

[4] 16-A.G. Taylor: **The organization of information – 2nd edition**
Westport Libraries Unlimited.

[5] 14-C. Shirky: **Ontologies are overrated: categories, links, and tags**
Clay Shirky's writings about the Internet
Retrieved from: **http://www.shirky.com/writings/ontology_overrated.html**

[6] 17-J. Trant: **Exploring the potential for social tagging and folksonomy in art museums: Proof of concept**
In New Review in Hypermedia and Multimedia, Volume 12 (1), June 2006 , pp. 83-105.

6. Web: **http://www.multimedian.nl/**.

[7] 21-H. Wu, M. Zubair and K. Maly: **Harvesting social knowledge from folksonomies**
In Proceedings of the seventeenth conference on Hypertext and hypermedia (Odense, Denmark, 2006), pp. 111-114.

[8] 15-M.A. Storey, L. Cheng, I. Bull and P. Rigby: **Waypointing and social tagging to support program navigation**
In Proceedings of the 2006 20th anniversary conference on Computer-supported cooperative work (Banff, Alberta, Canada, 2006), pp. 195-198.

[9] 23-J. Yew, F. Gibson and S. Teasley: **Learning by tagging: group knowledge formation in a self-organizing learning community**
In Proceedings of the 7th international conference on Learning sciences (Bloomington, Indiana, 2006), pp. 1010-1011.

[10] 13-S. Sen, S.K. Lam, A.M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F.M. Harper and J. Riedl: **Tagging, communities, vocabulary, evolution**
In Proceedings of the 2006 20th anniversary conference on Computer-supported cooperative work, (Banff, Alberta, Canada), pp. 181-190.

[11] 4-S.A. Golder and B.A. Huberman: **Usage Patterns of Collaborative Tagging Systems**
In Journal of Information Science, 32 (2). pp. 198-208.

[12] 5-M. Guy and E. Tonkin: **Folksonomies: Tidying up tags?**
In D-Lib Magazine, January 2006, Volume 12 (1)
Retrieved from: **http://www.dlib.org/dlib/january06/guy/01guy.html**.

[13] 18-K. Van der Sluijs and G. Houben: **Relating User Tags to Ontological Information**
In Proceedings of the Workshop on Ubiquitous User Modeling (UbiqUM) of IUI 2008 (Gran Canaria, Spain, January 13 - 16, 2008).

[14] 19-M. Van Setten and Ch. Wartena: **Using recommender system techniques to help users annotate content with tags** *[in preparation]*.

[15] 10-M. Melenhorst and M. van Setten: **Usefulness of tags**
In Proceedings of IEEE Professional Communication Society Conference (IPCC 2007), (Seattle, USA, September 30 - October 3, 2007).

Other reading

[16] 1-M. Ames and M. Naaman: **Why we tag: motivations for annotation in mobile and online media**
In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI 2007), (San Jose, USA, April 28 - May 3, 2007), pp. 971-980.

[17] 2-C.H. Brooks and N. Montanez: **Improved Annotation of the Blogosphere via Autotagging and Hierarchical Clustering**
In Proceedings of the 15th Worldwide Web Conference (WWW06) (Edinburgh, Scotland, May 23 - 26, 2006), pp. 625-632.

[18] 6-U. Mejias: **Tag literacy**
Retrieved from: **http://blog.ulisesmejias.com/2005/04/26/tag-literacy/**.

[19] 8-C. Marlow, M. Naaman, D. Boyd and M. Davis: **HT06, tagging paper, taxonomy, Flickr, academic article, to read**
In Proceedings of the seventeenth conference on Hypertext and hypermedia (Odense, Denmark, 2006), pp. 31-40.

[20] 11-P.C. Murray, Endeca Profind (2004): **Combining enterprise search and faceted classification**

In The Barrington Report on Advanced Knowledge Organization and Retrieval, I:1, pp. 19-27
Retrieved from: **http://www.kmconnection.com/Access_to_PDFs.htm**.

[21] 12-P. Schmitz: **Inducing ontology from Flickr tags**
In Proceedings of the WWW 2006, (Edinburgh, UK, May 22-26, 2006)
Retrieved from: **http://www.semanticmetadata.net/hosted/taggingws-www2006-files/22.pdf**

[22] 20-T. Wilson: **The strict faceted classification model**
Retrieved from: **http://facetmap.com/pub/strict_faceted_classification.pdf**

[23] 22-B.S. Wynar and A.G. Taylor (eds.), 1992: **Introduction to Cataloguing and Classification – 8th edition**
Libraries Unlimited.