Objective and Subjective Quality Assessment of Geometry Compression of Reconstructed 3D Humans in a 3D virtual room

Rufael Mekuria*[a], Pablo Cesar[a], Ioannis Doumanis[b], Antonella Frisiello[c]
[a]Centrum Wiskunde & Informatica, Science Park 123, Amsterdam, Netherlands, 1098XG;
[b]CTVC London. 9-10 Copper Row, Tower Bridge Piazza, London SE1 2LH,
[c]Instituto Superiore Mario Boella. Via P.C. Boggio 61 10138 Turin, Italy

ABSTRACT

Compression of 3D object based video is relevant for 3D Immersive applications. Nevertheless, the perceptual aspects of the degradation introduced by codecs for meshes and point clouds are not well understood. In this paper we evaluate the subjective and objective degradations introduced by such codecs in a state of art 3D immersive virtual room. In the 3D immersive virtual room, users are captured with multiple cameras, and their surfaces are reconstructed as photorealistic colored/textured 3D meshes or point clouds. To test the perceptual effect of compression and transmission, we render degraded versions with different frame rates in different contexts (near/far) in the scene. A quantitative subjective study with 16 users shows that negligible distortion of decoded surfaces compared to the original reconstructions can be achieved in the 3D virtual room. In addition, a qualitative task based analysis in a full prototype field trial shows increased presence, emotion, user and state recognition of the reconstructed 3D Human representation compared to animated computer avatars.
Keywords: Virtual World, Geometry Compression, objective quality, subjective quality

1. INTRODUCTION

Geometry compression has been studied actively in the scientific literature in the last decades. Techniques have been developed that can compress 3D polygonal meshes and 3D point clouds. These formats can be used to represent 3D objects of any topology. However, compared to image and video compression technologies, deployment of geometry compression is much less visibly present in existing IT infrastructures. Two key issues have hampered the widespread adoption of such technologies. First, much of the 3D content has been synthetically authored using computer software (CAD, 3D Modelling and animation software), resulting in relatively sparse data that can be stored and transmitted relatively easily without geometry compression. Second, contrary to image and audio data, intellectual property concerns seem to have hampered the widespread exchange of such 3D data in the internet. Nevertheless, with the current increase in technology for 3D scanning and 3D printing, the need for 3D geometry compression and standardized data formats is increasing. In the case of live scanning, the geometry is often much more dense (i.e. millions of points), and the rate of acquisition much higher (>10fps). This results in a much larger challenge of geometry compression. In addition, intellectual property issues seem less of a concern as data is much easier to acquire using 3D scanners. Therefore, these technologies require a renewed look at geometry compression, including its objective and subjective evaluation.

Fig. 1 Reverie Immersive virtual room, naturalistic user geometry is compressed and transmitted directly into the synthetic 3D scene

*r.n.mekuria@cwi.nl; phone +31(0)20 592 4020cwi.nl

Fig 1. shows the Reverie 3D immersive virtual room based on the Reverie framework [1]. In this application, reconstructed (scanned) 3D content is acquired and transmitted in real-time. It can be used as a realistic remote user representation. It shows that the acquired 3D user data is rendered remotely and compositely with the synthetic 3D scene. The remote transmission of such 3D geometry poses a large challenge in compression. Therefore, the Reverie system contains a compression framework for geometry compression to address the challenge of immersive communications. The framework currently includes 3 different codecs that have different properties and can be used in different situations depending on the need of the application. However, to use these codecs properly, an objective and subjective evaluation of their subjective quality and performance is needed. Therefore, the aim of this paper is to evaluate the quality of reconstruction and compression of 3D Humans in a 3D virtual room, both objectively and subjectively according to the following requirements.

Objective Quality Assessment

Support a full 3D triangle mesh or point cloud representation: Does the codec support the compression of triangle mesh or point clouds that are reconstructed on the fly from 3D scanners?

Low encoder and decoder frame delay: Is the structural algorithmic and computational delay below a certain threshold such that real-time end-to-end communications is possible?

Flexible/Progressive I/O representation: Does the compressed format have a "progressive" format, such that from partial bit-streams a coarser geometry can be decoded?

Color/normal: Does the codec support efficient, i.e. low-bit rate coding of color and perhaps normal attributes?

Inter-Predictive Coding: Does the codec efficiently exploit redundancy between subsequent frames?

Adaptability: Is the codec able to change coding parameters to adapt to different network, system and user conditions? i.e. does it have fine grained control of bit-rate, complexity, redundancy etc.?

Robustness to information loss: Does the codec provide resilience to data losses?

Bandwidth: what is the resulting bit-rate and resulting distortion performance (R-D performance) of the codec?

Systems support and integration in overall framework. Can the codec be integrated easily in a mixed reality system, i.e. does it use generic API and programming interfaces?

Subjective Quality Assessment

Subjective degradation compared to the original reconstructed mesh: what is the perceived perceptual degradation compared to the originally reconstructed mesh?

Subjective experience compared to Computer based avatar: what is the user experience of having a natural reconstructed user, instead of a computer based avatar, i.e. such as those used in traditional virtual environments?

This goes a step beyond traditional quality evaluation of geometry compression, which generally only considers algorithmic complexity, rate-distortion and the memory footprint introduced by the codec. Instead, we present a full comparative quality evaluation of the codecs in the Reverie compression framework using the Reverie immersive virtual world. This evaluation includes both subjective and objective testing based on the criteria described above. In addition, the user experience of live reconstructed geometry compared to other 3D representations such as synthetic avatars in the 3D world is studied based on interviews and field trials of the Reverie system deployed with end users.

The paper is structured as follows, in section 3 we present the 3 codecs and their integration into the Reverie system. In section 4 we present the objective evaluation that includes qualitative comparison based on properties and a quantitative comparison that includes: rate-distortion and real-time implementation performance (latency). In section 5 we present the subjective results of the achieved quality in the immersive virtual room. In section 6 we conclude the paper and provide future recommendations on compression for immersive communications. We present some of the related work in the next section.

2. RELATED WORK

Over the years, both objective and subjective testing of geometry compression distortion has been studied. To measure distortion caused by mesh compression schemes objectively, often symmetric distance measures such as sym. root mean square (symm. rms) or sym. Haussdorf have been used. Objective tools that support efficient computation of these measurements have been presented in [2] and [3]. Additionally, works in [4] and [5] provide an additional overview of both subjective and objective testing of geometry quality and the factors involved (i.e. geometry quality vs. texture quality and so forth). However, none of these works address the objective and subjective quality of time varying geometry compression in an immersive virtual world. This paper studies integration of codecs in a mixed reality immersive system and presents the resulting objective and subjective performance results.

3. GEOMETRY COMPRESSION

MPEG-4 PART 16 AFX, TFAN profile

Standardization of compressed bit stream formats is of great importance, as it guarantees interoperability between devices, storage and network services. Standards for geometry compression have been developed in ISO/IEC JCT SC 29 WG11 also known as the motion picture experts group (MPEG). MPEG is the premier organization for standardization of media data types, such audio, video and 3D graphics content. The current standard for mesh geometry compression can be found as MPEG-4 part 16 Animated Frameworks Extension (AFX), which is part of MPEG-4 (ISO/IEC 14496). The algorithm behind the latest and most efficient compression profile TFAN was presented in [6]. In addition, there is an open source version of the reference software available at http://www.mymultimediaworld.com/. We explain the principles of the TFAN algorithm below. TFAN works by first decomposing the mesh into triangle fans, which are an ordered set of k triangles and k+2 vertices. Each of the ordered triangles inside a TFAN are neighbors, and the triangles have the same orientation and share a common vertex (the central vertex). The decomposition of the mesh in TFANs results in overlapping triangles between fans. In Fig. 2, we show an example TFAN as it is detected, in this case the blue triangles in the TFAN are "non-visited", while the purple ones are "visited" as they occurred in a previously coded fan. Key to TFAN is the detection and definition of 9 different configurations and their relative frequency of occurrence. Based on these relative frequencies, bits are efficiently allocated. Configurations are defined by the degree of the TFAN (number of triangles), the set of visited and non-visited vertices and a relative index vector (each TFAN uses a relative vector for indexing resulting in a lower number of bits for indexing). These operations result in a very compact representation of the connectivity. As a side effect, this representation introduces changes to the order of the triangles (which does not matter for rendering or further processing the mesh, but might matter in some very specific situations such as when correspondence between vertices in different meshes is needed). Subsequently, the connectivity information is used to compress the vertex coordinates and attributes via inter-vertex prediction based on Differential Pulse Code Modulation (DPCM) or parallelogram prediction (the latter uses three vertices to predict a fourth vertex). The resulting prediction residuals are compressed via Context Adaptive Binary Arithmetic Coding (CABAC) or other entropy coding schemes similar to those that are used in state of art video compression codecs.

Based on the open source implementation, a modified version has been integrated in the Reverie immersive framework to enable MPEG-4 based geometry compression.

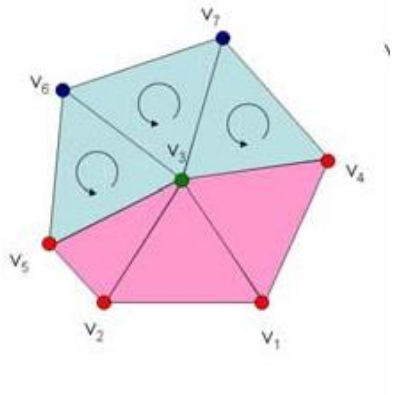

Fig. 2: An example of a common TFAN configuration in a (manifold) mesh

Connectivity Driven Geometry Compression for immersive virtual room

Alternative to the standard based approach based on TFAN, a geometry compression algorithm in [7] was proposed. This algorithm introduces a simpler/heuristic method for connectivity encoding, and coarse quantization of the differentials in order to reduce the encoding complexity. Key aim of the codec is to benefit from the properties of dense 3D reconstructed data as much as possible without introducing perceptual artifacts. This codec was developed with the aim of compression of dense geometry for a 3D immersive virtual world specifically in mind and aims to enable encoding real time encoding of dense reconstructed geometry.

The block diagram of this algorithm on the encoder side is shown in Fig 3. The input mesh is M(V,E), where V represents the vertex data and C(E) the connectivity data. The data passes 3 different parallel paths; the path in the lower part of Fig. 3 constitutes the connectivity encoder. The connectivity coder does a differential between indices in subsequent triangles in the list. In reconstructed data, subsequent triangles in a connectivity list often share vertex indices, or indices in a close range. This differential step therefore results in values in a smaller integer range, mostly centered around 0. This new differential connectivity data can therefore be stored using less bits and efficiently entropy encoded, in this case using zlib, an open source compression library for entropy encoding based on the deflate algorithm. In addition, a technique has been integrated to detect patterns in the connectivity, i.e. repeating regularities introduced from the reconstruction process. These patterns can be encoded in a run length fashion, by indicating the pattern and the number of repetitions. In case the triangulation of the 3D reconstruction introduces such patterns, an additional coding gain can be achieved. More details on the method on this technique can be found in [7] .

The upper part of the diagram in Fig. 3 illustrates the encoding of the geometry. In this case, it is based on a differential quantization of connected geometry positions. The connectivity is traversed and the differential is computed between connected vertices. This differential is then quantized coarsely, in order to save bits. The coarse non-linear 4 bits quantizer A is pre-trained on a narrow range, that constitutes over 90 % of the residuals, then for larger values, linear quantizers (A,B,C) can be used (8,16 and 32 bits respectively). The result of this is that connected geometry can be quantized efficiently and rapidly without large quantization artifacts. The middle stage of the codec introduces appearance quantization, which is a quantization scheme similar previously described, but for the normals and colors. The overall compression efficiency of this codec at comparable qualities compared to MPEG-4 TFAN is 10-15 % less on test/training data. However, it was able to run in real-time on live data in our immersive system, enabling real-time end-to-end communications. Key speed optimization was achieved by training the quantization tables beforehand for given reconstructed training data.

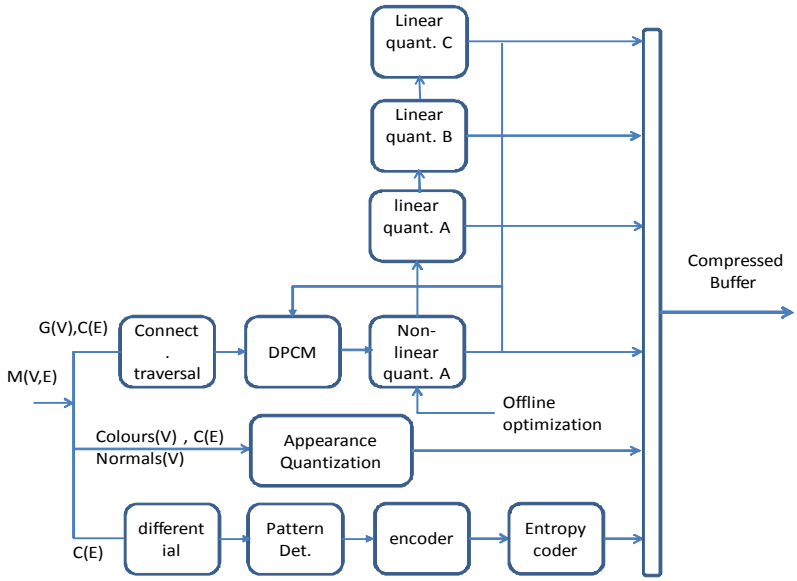

Fig. 3 Geometry encoder based on [7]

Geometry driven compression for immersive virtual room

Fig. 4 illustrates a geometry driven mesh coding scheme for the immersive virtual room. Contrary to the approach presented in the previous section, this algorithm is based on first coding the geometry information (V) using an octree composition scheme, and only after that code the connectivity information. The octree composition results in a voxel space composition, denoted $v_{octree}\{x,y,z\}$, where x,y,z is a key that can be used as an index in the voxel grid.

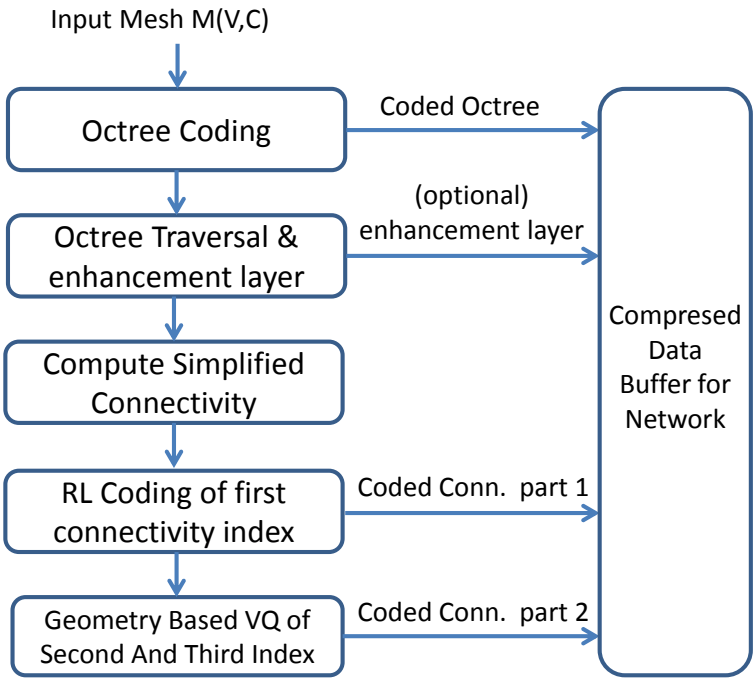

Fig. 4: Geometry Driven Mesh Codec based on [8]

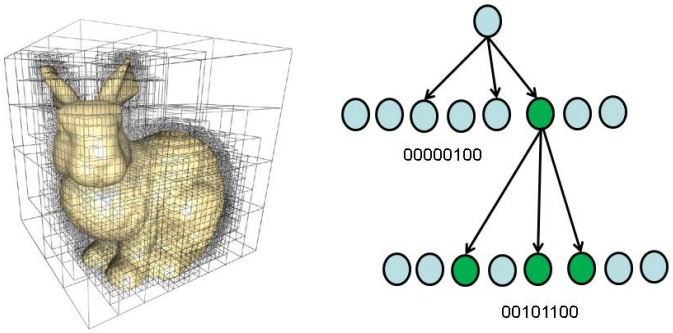

Fig. 5 Octree subdivision in space

In the second stage, we traverse the octree grid v_{octree} storing the indices I_{simp} which are single integer indices to voxels. I_{simp} contains all vertex indices for the new simplified connectivity $C_{simplified}$, which will be derived from the original connectivity $\boldsymbol{C}_{\text{original}}$ of the mesh. During this traversal, we optionally compute an enhancement layer that refines the position of the voxel center based on the positions of enclosed vertices based on a weighted average. This layer can be used to increase in the decoded mesh quality without further octree subdivision. This traversal gives two mappings that are useful for coding connectivity information. First, the mapping m from the original vertex indices I_{or} to the simplified leaf voxel indices I_{simp} and, second the mapping l between I_{simp} and the 3D voxel indices $v_{octree}\{x,y,z\}$ in the grid V_{octree}.

$$m: I_{or} \rightarrow I_{simplified}\ \{i_{or} \in I_{or}, m(i_{or}) \in I_{simplified}\} \quad (1)$$

$$l : I_{simplified} \rightarrow V_{octree}\{i_{simp} \in I_{simplified}, l(i_{simp}) \in V_{octree}\} \tag{2}$$

We present the original connectivity $C_{original}$ as a collection $T_{un\_ordered}$ of triples of unordered indices and the new simplified connectivity as $C_{simplified.}$ as a collection $T_{ordered}$ of index triples in ascending order. $T_{ordered}$ and $T_{un\_ordered}$ represent collection of indix triples as given by relations (3) and (4) below. $C_{simplified}$ is then the collection $T_{ordered}$ that satisfies relation (5).

$$T_{unordered} = \{i_1, i_2, i_3\} : \{i_1, i_2, i_3 \in I_{original}, i_1 \neq i_2 \neq i_3\} \tag{3}$$

$$T_{ordered} = \{i_1, i_2, i_3\} : \{i_1, i_2, i_3 \in I_{simplified}, i_1 < i_2 < i_3\} \tag{4}$$

$$\begin{aligned} C_{simplified} = \{&t_{ordered}(i_1, i_2, i_3) \mid \exists tunord(i_{or1}, i_{or2}, i_{or3}) : \\ &t_{unord} \in C_{original}, m(i_{or1}) = i_1, m(i_{or2}) = i_2, m(i_{or3}) = i_3\} \end{aligned} \tag{5}$$

Ordering of the vertex indices i_{simp} as $t_{ordered}$ in $C_{simplified}$ helps to preserve more of the spatial correlation introduced by the structured octree traversal and detect duplicate triples. Next, the set $C_{simplified}$ is ordered in ascending order of i_1 and all i_1 are coded via a run length coding scheme, that codes the number of repetitions and increments in i_1. This is part is the run length (RL) coding method for the first connectivity index in Fig. 4. Next, we construct geometry based representations t_{VQ} of each triangle $t_{ordered}$ in $C_{simplified}$ to code the second and third indices i_2, i_3. Every t_{VQ} represents a $t_{ordered}$ where i_2 , i_3 are represented by a 3D offset in the octree voxel grid V_{octree} away from connected voxels $v_1=l(i_1)$ and $v_2=l(i_2)$. Hence the elements of the set T_{VQ} satisfy relation (6):

$$T_{VQ} = \{i_1, z_2, z_3\} : \{i_1, \in I_{simplified}, z_2, \in Z^3, z_3, \in Z^3\} \tag{6}$$

Z_2 and z_3 are signed discrete 3D vectors representing a shift in the octree voxel grid V_{octree}. The T_{vq} representations are obtained from $T_{\text{ordered.}}$ We describe this by the mapping F_{vq} given in (7):

$$F_{vq} : T_{ordered} \rightarrow T_{VQ} = \{i_1, l(i_2) - l(i_1), l(i_3) - l(i_2)\} \tag{7}$$

In each t_{vq} in $T_{vq,}$ i_1 was already coded by the RL coding of the first index and therefore, only z_2 and z_3 need to be encoded. By structuring the connectivity information in $T_{ordered}$ and F_{vq} we achieved a structure where z_2 and z_3 can be coded very efficiently via vector quantization. The prototype vectors *[-1,0,0], [0,-1,0]* and *[0,0,-1]* occur over 75% of the cases in our training data for z_2 and z_3 while around 15% of the other cases are covered by the vectors *[-1,1,0], [0,-1,1], [1,-1,0].* The remaining signed binary vectors represent the last 7% of the cases in our training data. We developed an efficient variable length coding (VLC) scheme to code these vectors with 2, 4 and 8-bit code words. In the exceptional other cases, we store all components of z_2 and/or z_3. The decoder executes inverse operations, i.e. the octree voxel structure is decoded, and instead of only l we also compute the inverse l^{-1} that relates the 3D octree voxel index v_{octree} to $I_{simplified}$, i.e.:

$$l^{-1} : V_{octree} \rightarrow I_{simplified}\{v_{octree} \in V_{octree}, l^{-1}(v_{octree}) \in I_{simplified}\} \tag{8}$$

The mapping l^{-1} is used to decode all $t_{ordered}$ recovering $C_{simplified}$ from all t_{vq} using the mapping F^{-1}_{vq} *:*

$$F^{-1}{}_{vq} : T_{VQ} \rightarrow T_{ordered} = \{i_1, l^{-1}(l(i_1) + z_2), l^{-1}(l(i_2) + z_3)\} \tag{9}$$

Where i_1 was already recovered from the run length encoded data and therefore, by recovering i_2 followed by i_3 for each t_{vq}, $C_{simplified}$ is recovered. By having the geometry and both connectivity, the mesh is fully decoded. Therefore, we have presented a complete lossy geometry compression scheme that allows very fine grained control of the output resolution (number of output points) by changing the octree depth. In addition, the codec enables real-time encoding and decoding, useful for real-time end-to-end communications.

For more information on the codec algorithm and its Rate distortion performance we refer to [8].

Integration in Immersive Virtual Room

We provide implementation details of the integration in the immersive virtual room in Fig. 6 and Fig**. 7**. Fig. 6 outlines the run-time code. A 3D reconstruction source in a capture thread generates the 3D Mesh data. When this data is reconstructed, a compression thread is triggered via an event and the new frame is compressed (or queued when this thread was still busy encoding). The compression thread runs the encoder and writes data to memory shared with the network thread, which is responsible for the transmission. The implementation at the receiver is analogous. The parallel pipelined execution of capture, compression and networking enables us to further reduce the end-to-end delay and achieve higher frame-rates. While this configuration is quite similar to that seen in video conferencing systems, there is one issue that can hamper the performance. Contrary to when such a configuration is deployed in video conferencing systems, frames can have very different sizes (due to the variable number of points in the mesh, which don't have a fixed resolution). This can result in highly varying bit-rate/computational loads, resulting in the capture/compression/network threads running out of sync. This in turn, can result in having an excess amount of frames in the pipeline and long end to end delay. To combat this, a first in first out (FIFO) scheme was adopted in the frame transmission. In this scheme, only the newest (latest) mesh frames are compressed, transmitted or rendered, while late frames are discarded. This minimizes the end-to-end latency of frames.

In Fig. 7 we illustrate the class integration of the compression framework. The base class Nano3DCompression provides the base encoder and decoder function for meshes in the Nano3D format, which is a customized mesh format uniformly used throughout the Reverie framework. The base class has a static factory function to create instances based of the derived classes, which implement the compression methods described in section 3. The Nano3DSMC contains the MPEG-4 codecs, the Nano3DIMTPC the connectivity driven method and the Nano3DGeom the geometry driven coding scheme. Based on this integration different compression methods can easily be used and integrated without changes to the capture run time code. This enables easy experimentation and easy deployment of different codecs.

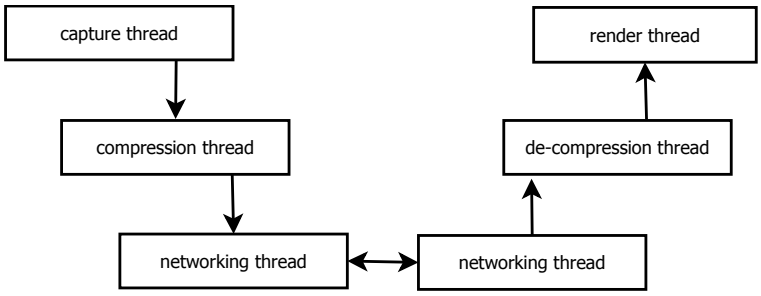

Fig. 6 Threading of geometry compression for remote transmission

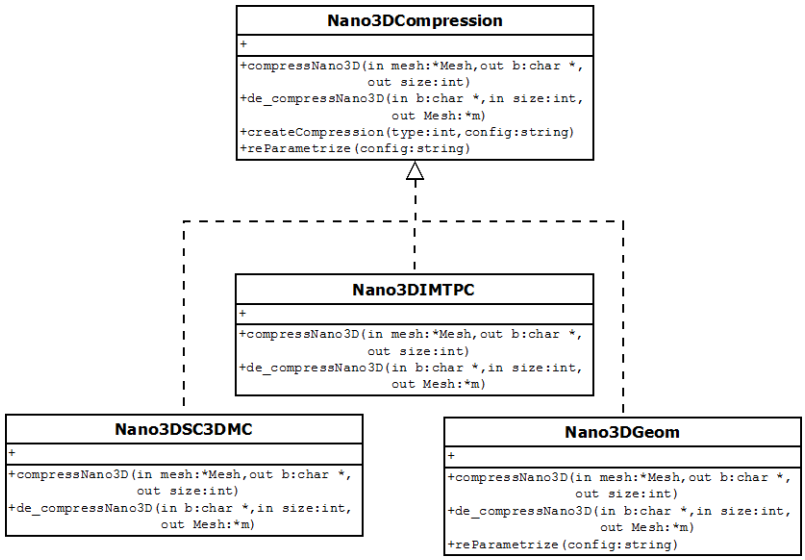

Fig. 7 Integration of compression functions

4. OBJECTIVE QUALITY ASSESSMENT

Comparative evaluation of codecs in the Reverie framework

Table 1 Comparative Evaluation of properties of codecs integrated in the Reverie framework

Codec	Mesh	Point Cloud	Low Delay	Flexible I/O	Color	Normal	Inter-Predictive	Adaptability	Information loss	Rate / Distortion
MPEG-4 TFAN	Y	N	+	+	Y	Y	N	+	--	++
Conn Driven	Y	N	++	-	Y	Y	N	--	--	+
Geometry Driven	Y	Y	+	++	Y	N	N	++	+	+

In Table 1 we compare the codecs integrated in the framework based on properties as described in section 1. All codecs are able to directly compress the 3D mesh geometry resulting from 3D surface reconstruction. The geometry driven codec can also compress point clouds (i.e. point sampled data without the connectivity information). The MPEG-4 and connectivity driven codecs need the connectivity to code the geometry and can therefore not be used to encode point clouds directly. All codecs have low delay properties, but in different ways. The MPEG-4 TFAN Codec has a linear encoding and decoding time and can decode in real-time. For dense meshes with over 300K points (as present the test and training data) however, we were not able to run the encoder in real-time on our hardware/software setup (i7, 3.400 GhZ desktop machines, 16GB Ram, Windows 7, VS 2010). However, for meshes with vertices in the range of 10K points the encoder can also run in real-time (<200ms delay) in our hardware/software setup. The integrated connectivity driven codec is the fastest codec, especially when used with dense meshes, as it is mostly based on computing differentials and quantization only and provides real-time encoding. All codecs provide coding of the colors and MPEG-4 TFAN and the connectivity driven codec can also encode normal information. We are not sure if it is preferable to encode the normals or instead re-compute them at the receiver. This is a design choice, and we had good results in both cases. The codecs currently do not support inter-predictive coding. Inter predictive coding of time varying point clouds and 3D mesh geometry has not been studied well in the literature. Therefore, developing appropriate techniques that run in real time and can be integrated in the immersive framework is beyond the scope of this paper and the Reverie framework. In terms of adaptability, i.e. the degree in which one is able to tune the coding complexity, i.e. gain, complexity etc., the geometry driven codec is best. The geometry driven codec can be used to decrease the number of output vertices compared to the input, which enables fine grained quality control and low bit-rate encoding. This feature is not present in MPEG-4 TFAN and the connectivity driven encoder. The MPEG-4 TFAN Codec only allows quantization bits for the geometry, normal and color data components to be specified. This enables some tuning of the bit-rate, but for real time encoding of dense meshes at low bit rates and in variable network conditions this is quite limited. In practice, reducing the amount of points in the geometry driven codec is a better way to achieve a lower quality representation than direct quantization without vertex decimation. The connectivity driven codec has no support to change the bit-rate settings on the fly, which is its key disadvantage. Last, both the MPEG-4 TFAN and the connectivity driven codec do not provide resilience to data/information loss. In current networks, such information losses are often corrected at lower layers, so it tends not to be a very big problem. The geometry driven codec does not provide full error resilience, but instead provides a progressive format. This implies that from a partial stream a lower quality object can be obtained. This facilitates context adaptive forwarding of data and application layer routing based on available bandwidth characteristics and application needs. In terms of rate-distortion, the MPEG-4 TFAN codec provides the best rate-distortion characteristics, while the geometry and connectivity driven codecs are slightly worse in this respect. On the other hand the last two provide better support for real-time and low bit –rate encoding in the immersive framework, and are used in the study in section 5.

Geometry Quality and Bit-Rate Assessment

In Fig. 8 we show the achieved geometric quality versus the byte size per encoded frame on reverie test data reconstructed with the system in [9]. The graph plots bit-rate on the horizontal versus distortion on the vertical axis. In such comparison, the lower the curve, the better the codec is in the rate-distortion sense. The purple line represents the geometry driven codec at different settings. For MPEG-4 TFAN we present the rate-distortion for two settings (8 bits

and 10 bits per geometry component, and with 6 bits per color/normal component), in the red and green dots. The connectivity driven codec is represented in the preconfigured setting (the blue dot). We measure the distortion on the vertical axis using a symmetric root mean square (rms) distance measure between original and decoded models using the metro tool [3], we refer to [3] for more details on this metric. What can be seen is that the connectivity driven codec and MPEG-4 TFAN are better in the rate distortion sense, but that the geometry driven codec can provides more configurations for low bit-rate encoding. This is very useful in the virtual room environment as these different properties complement each other, highlighting the benefit of having multiple codecs in the framework. We can use the geometry driven codec when a low bit-rate and low quality is needed, and gradually increase the quality. For high quality representations we can use the connectivity driven coder for dense geometry or MPEG-4 TFAN in case meshes are sparse. In all these cases, real-time encoding and decoding can be maintained, enabling real-time end-to-end communications in the Reverie framework.

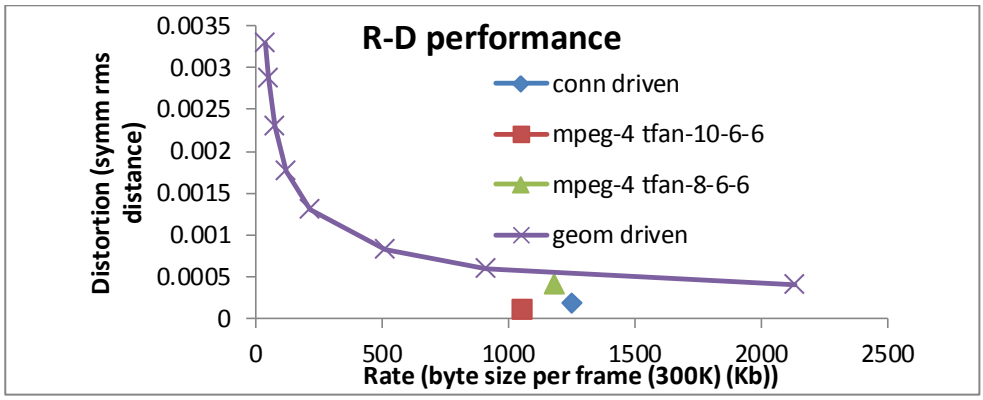

Fig. 8 Rate-distortion performance codec in the immersive virtual room

Real-Time Performance Assessment

In Fig. 9 we assess the real-time performance of the different codecs running in the integrated immersive communications framework. The numbered entries represent different settings of the geometry driven encoder. The settings for the connectivity driven and MPEG-4 TFAN correspond to those in the previous section. The red value corresponds to the decoder time, while the blue line represents the encoder time. We can see that the encoder/decoder time increases exponentially for higher quality settings the geometry driven codec. For higher qualities we should therefore switch to the connectivity driven codec to still enable real-time encoding in the immersive framework. The MPEG-4 TFAN was mainly designed with offline encoding in mind, which is observed in Fig. 9. However, MPEG-4 TFAN can be used for encoding lower resolution meshes complemented with textures. Again, this comparison illustrates the complementary nature of the different codecs integrated in the framework. In the Reverie framework, geometry driven codec is used for low bit-rate real-time encoding, the connectivity driven codec for high bit-rate real-time and TFAN for low res input meshes with textures. This also suits very well the different 3D reconstruction modules integrated in the Reverie framework such as [9] (geometry/connectivity driven) and [10] (based on MPEG-4 and JPEG for textures coding.

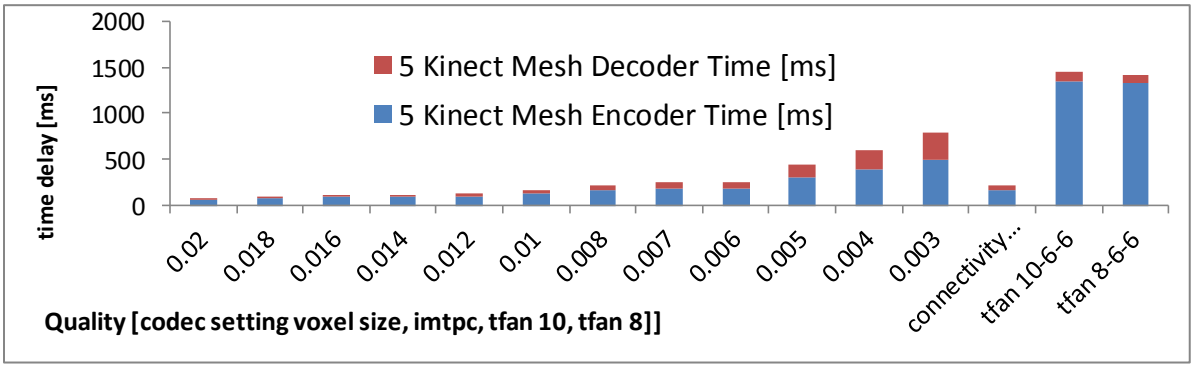

Fig. 9 Real-Time performance on reconstructed mesh data with over 300K vertices

5. SUBJECTIVE QUALITY ASSESSMENT

Quantitative Study on Geometry Quality

A laboratory test has been performed according the procedures suggested in ITU Standards [11], [12]. 16 volunteer representative users have been recruited. The participant sample population was gender balanced, the age range from 25 to 38 years old (mean 30,8), with a high educational level (degree, master or PhD), mainly in engineering branches. The users provided their informed consent to participate to the experiment when invited. The user study was previously presented in the streaming framework developed in [13].

A within test has been designed: in order to estimate how individual behaviour changes when the stimuli and variables change, each subject has been asked to assess all the configurations to be tested [14]. Two iterations have been performed with different subjects. The stimuli to evaluate were a sequence of randomized videos, presenting the same subject, a 3D human moving at 8 fps in a virtual environment created within the Reverie system (hangout virtual room) that includes a high quality 3D rendering and degradation introduced by the compression method. The independent variables observed were:

- The compression methods: original, connectivity driven [7], geometry driven [8] (parameter 0.008), and enhanced
- The virtual distance of the 3D reconstructed user (far vs close to the virtual camera in the 3D space)

The subjects have been welcomed in the laboratory, and before starting the test received the introduction to the experimental goals, highlighting the focus on the 3D human and the context of the study. The possibility to interrupt the test if needed was provided. Before evaluating the videos, each subject had the possibility to interact with the real system to become more familiar with the context of realistic users in a 3D virtual room. Then, they were asked to watch pre-recorded stimuli videos one at a time (in a randomized order). After each video the subjects expressed an evaluation of the perceived video quality of the 3D reconstructed human. Additional rating dimension of the perceived realism into the scene (explained as the integration of the 3D human render with the virtual environment) was given.

Each dimension has been evaluated by using the MOS – Mean Opinion Score [15], allowing to collect the average of the opinions ("votes") on the single given conditions. The rating scale (on 5 points: Excellent 5, Good 4, Fair 3, Poor 2 , Bad 1) was printed and left on the table as reference for helping the subject to reply. The test protocol applied allows collection of the subjective evaluations on the single video just after it is watched, on the basis of the immediate impression. For each test variable, the mean value and the standard deviation of the statistical distribution of the assessment grades has been calculated. Table 2 shows modest differences between original and decoded meshes. Pairwise comparison of each codec setting with the original mesh via paired samples t-tests, showed that only for the geometry driven codec (0.008) [8] the mean difference was significant ($p<0.05$). The paired sampled t-test is a statistical test to compare the means of two different normally distributed variables. To check its validity normality assumptions have been asserted. Next, we investigate the effect of nearby and far away camera views in the 3D word. For geometry driven both near and far were tested significantly lower compared to the original mesh ($p<0.05$). On the other hand, differences of the original with connectivity driven were not found significant.

Table 2 – MOS on the perceived quality

	MOS and Compression method							
	Original [9]		Conn driven [7]		Geometry driven [8]		Geoemtry Driven enhanced [8]	
	mean	st.dev	mean	st.dev	mean	st.dev	mean	st.dev
quality	3,2	0,9	3,0	1,0	2,7	0,9	2,9	1,0
realism	3,3	0,8	2,9	0,9	3,0	0,9	2,9	1,0

Table 3 MOS at different virtual distances

	MOS and Virtual distance			
	Near		Far	
	mean	st.dev	mean	st.dev
quality	2,8	0,9	3,1	1,0
realism	2,9	0,9	3,2	0,9

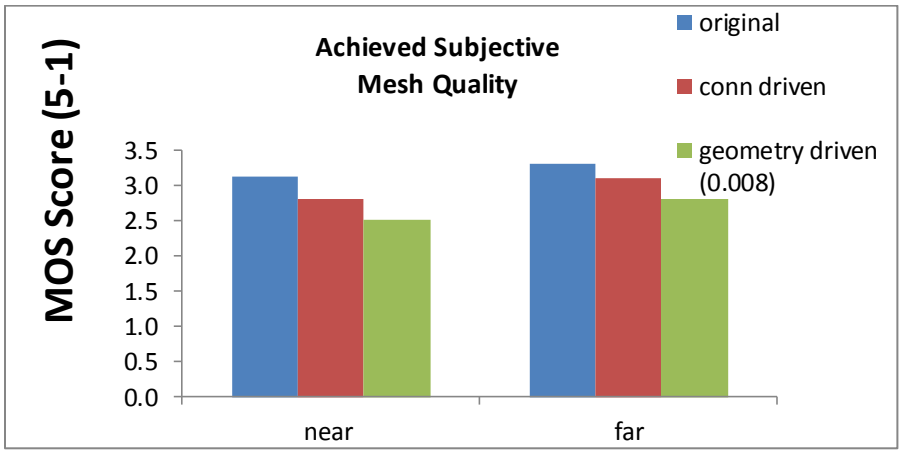

Fig. 10 Subjective results evaluation natural user quality (original, conn driven, geometry driven)

Qualitative Study on the usefulness of reconstructed 3D user compared with Synthetic Avatars

The Reverie immersive system contains a hangout scenario where 2 users are represented as synthetic computer avatars (referred to as avatar representation) and 2 as reconstructed 3D Mesh users (referred to as replicant representation). The 4 users interact in a virtual room. This scenario was deployed in a field trial to evaluate the user experience. This trial allowed us to compare the user experience of synthetic avatar users to that of reconstructed 3D Humans. More information about these field trials can be found in [16]. In this study, participants in groups of four (two as replicants and two as avatars interact in the virtual room to complete a collaborative task. In this scenario, one of the replicant users was given a step-by-step manual on how to create two objects using the Lego Mega Blocks. This user had to communicate how to build the shapes using both verbal and non-verbal means to the rest of the group (avatars and replicant). The group had to replicate the shapes on a notepad using words to describe their various features (e.g., color and shape). The analysis of the qualitative data of the reconstructed 3D users showed the following remarkable difference between avatar and reconstructed 3D human:

The replicant representation improves the following four concerns of the avatar representation in the virtual world.

1) *Presence:* The user knows if who and if another user is present in the virtual world, contrary to the avatar representation where this is not immediately clear.
2) *Physical recognition:* The replicant user is instantly recognized and is seen as an excellent representation of the actual person, much more so than the avatar representation.
3) *Current user state:* The user can instantly recognize what the user does and where s/he looking at. For example, participants recognized the usefulness of the replicant representation in the successful completion of the collaborative task.
4) *Emotional state:* The user can instantly see the user's emotional state without the need to simulate any facial/body movement.

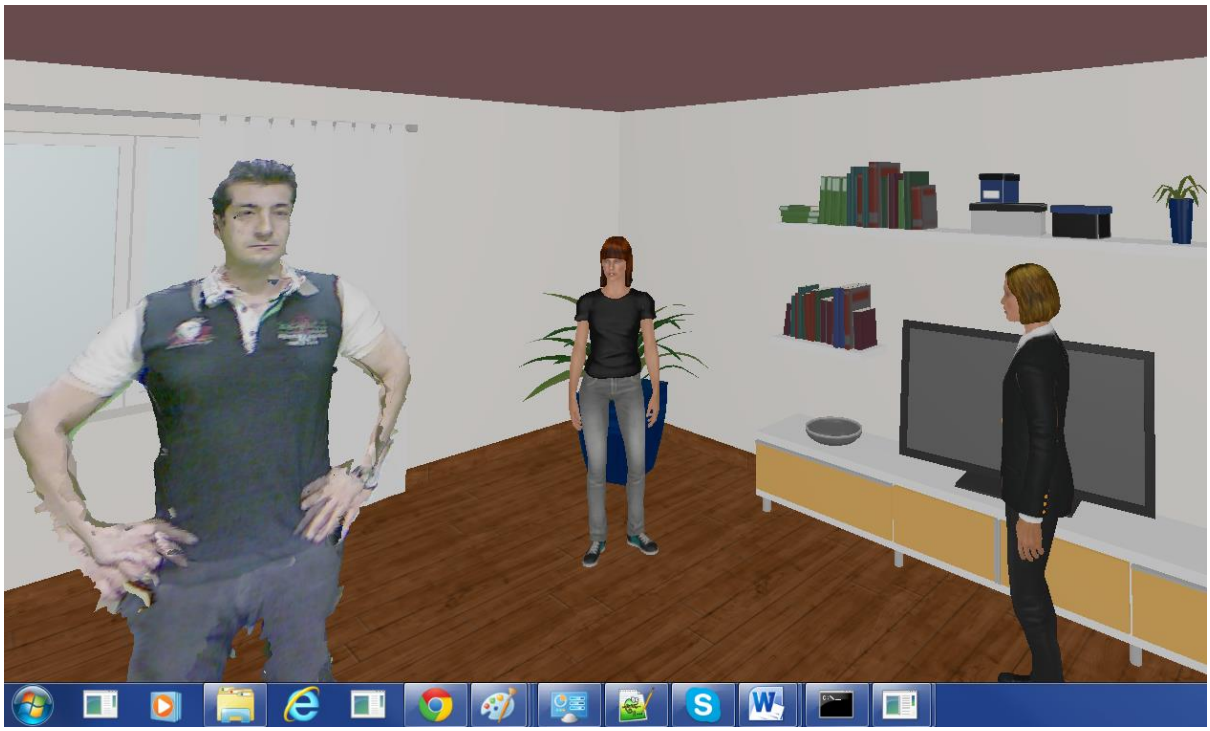

Fig. 11 Reverie hangout Scenario, natural replicant user interacts with synthetic avatar user

Contrary, the avatar representation showed two key advantages compared to the replicant representation

1) *Odd feeling:* Low-quality replicants (e.g., with missing user body parts or with artifacts) degrade the user experience. If the human user is not fully and accurately reconstructed (e.g., without any artifacts around the human figure or any missing body parts), the replicant representation creates an "odd" feeling to all participants in the virtual world. Furthermore, completing tasks with low-quality replicants, require users to rely upon verbal channels only and to ignore the non-verbal ones to successfully complete a task. While most such artifacts are introduced by 3D reconstruction, it is important that the codec does not introduce such artifacts.

2) *Fixed Identity:* Replicants are not useful for users who want to adopt a different identity online. Regardless of the quality of reconstruction, replicants are instantly recognizable by the users in the virtual environment. This makes the replicant representation unsuitable for users who do not want to reveal their identity online. For such users, the avatar representation is the preferred alternative. The avatar representation enables users to hide their identity while maintaining a high-level of control over their virtual representation. This means that users are still able to embody their avatars in a "natural" way by pose and gesture and emotion recognition.

The results of these field trials highlight the great potential of the replicant representation in virtual worlds, and the relevance of design, implementation and evaluation of geometry codecs to support this representation.

6. CONCLUSIONS

In this paper we evaluated the objective and subjective merits of a set of geometry compression algorithms in a 3D immersive virtual environment. The geometry codecs represent typical mesh coding paradigms *connectivity driven* and *geometry driven* and complement each other well to support real-time immersive communications in a virtual room. The quantitative subjective study has shown that the codecs can introduce negligible distortion compared to the original 3D reconstructed geometry when the right codec is chosen (connectivity driven for nearby and geometry driven for far away). In addition, a field trial experiment in the hangout virtual room scenarios showed that the replicant representation increases presence, physical, emotional and user state recognition compared to synthetic computer avatar users

7. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ICT-2011-7-287723 (REVERIE project). We thank all the partners in REVERIE project for their useful contributions to the project. We thank Kees Blom for proof reading this work. We thank Khaled Mamou and Insititut Mines Telecom for open source MPEG-4 TFAN implementation.

8. BIBLIOGRAPHY

[1] Reverie FP7. Reverie. [Online]. http://www.reveriefp7.eu/

[2] Aspert N, Santa-Cruz D., Ebrahimi T., "MESH : Measuring Errors Between Surface Using The Haussdorf Distance," in *IEEE International Conference in Multimedia and Expo*, Laussane, CH, 2002, pp. 705-708.

[3] Cignoni P. , Rocchini C. and Scopigno R., "Metro: measuring error on simplified surfaces," *Computer Graphics Forum*, vol. vol. 17, no. 2, pp. 167-174, June 1998.

[4] Bulbul A., Capin T., Lavoue G., Preda M., "Assessing Visual Quality of 3-D Polygonal Models," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 80-90, Nov. 2011 2012.

[5] Berjón D., Morán F., Manjunatha S., "Objective and subjective evaluation of static 3D mesh compression," *Signal Processing: Image Communication*, pp. 181-195, 2013.

[6] Mamou K., Prêteux F., Zaharia T. "TFAN: A low complexity 3D Mesh Compression Algorithm," *Computer Animation and Virtual Worlds*, vol. 20, no. 12, pp. 343-354, June 2009.

[7] Mekuria R. and Bulterman D., Cesar P., "Low complexity connectivity driven dynamic geometry compression for 3D Tele-Immersion," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, Florence, 2014, pp. 6162,6166.

[8] Mekuria, R., Cesar P. "A Basic Geometry Driven Mesh Coding Scheme with Surface Simplification for 3DTI.," *MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE e-letter*, vol. 9, no. 3, pp. 6-7, May 2014.

[9] Alexiadis D., Zarpalas D., and Daras P.,., "Real-Time, full 3-D reconstruction of moving foreground objects from multiple consumer depth cameras," *IEEE Transactions on Multimedia*, vol. 15, pp. 339-358, 2013.

[10] Alexiadis D., Zarpalas D., and Daras P., "Real-time, realistic full-body 3D reconstruction and texture mapping from multiple Kinects," in *IEEE 11th IVMSP Workshop, 2013*, Seoul, 2013, pp. 1,4.

[11] ITU-T , "Methods for objective and P.800.2 subjective aassessment of speech quality. Mean opinion score interpretation and reporting," ITU Recommendation P.800.2, 2013.

[12] ITU-T, "P.910 Audiovisual quality in multimedia services. Subjective video quality assesment methods for multimedia appilcations," P.910,.

[13] Mekuria R., Frisiello A., Pasin M., and Cesar, P., "Network support for social 3-D immersive tele-presence with highly realistic natural and synthetic avatar users," in *7th ACM International Workshop on Massively Multiuser Virtual Environments (MMVE '15)*, Portland, OR, 2015, pp. 19-24.

[14] Charness G., Greezy G., Kuhnc ., "Experimental methods: Between-subject and within-subject design," *Journal of Economic Behaviour & Organization*, pp. 1-8, Aug. 2012.

[15] ITU-T, "Methods for objective and subjective assessment of speech quality. Mean opinion score interpretation and reporting," ITU Recommendation P.800.2, 2013.

[16] Reverie Consortium, "D 3.6. Report on Reverie Field Trial," EU FP7 Deliverable 2015.