

Plan détaillé de relecture "Audio 3D"

TE5690

Version préliminaire - Ne pas diffuser

Médias immersifs : vidéo 360° & audio 3D
- Principes. Systèmes, techniques, normes

*Immersive media: 360 video & 3D audio -
Basics. Systems, techniques, standards*

Jean-Noël GOUYET

Ingénieur-formateur et consultant en techniques et gestion des médias numériques.
Ancien chargé d'études à la Direction de la Recherche de l'INA.

Résumé

Cet article est le premier d'une série de trois, concernant les systèmes audio-visuels immersifs. Il en définit d'abord les composantes : le contenu omnidirectionnel (vidéo à 360° et audio 3D), le continuum du réel au virtuel et ses divers degrés (*Augmented Reality, Mixed Reality, Augmented Virtuality, Virtual Reality*) avec les facteurs d'immersion et d'interactivité, les interfaces-utilisateurs (visiocasque, binaural...). Dans une deuxième partie il détaille les fonctions et les techniques : l'acquisition, le traitement, le codage et formatage, la distribution et l'utilisation. La dernière partie précise les normes de codage et de formatage pour la distribution, en particulier celles de l'ensemble *MPEG-I Immersive Media* (OMAF...).

Abstract

This article is the first of a series of three about immersive audio-visual systems. It first defines their components: the omnidirectional content (360° video and 3D audio), the continuum from reality to virtuality and its various degrees (Augmented Reality, Mixed Reality, Augmented Virtuality, Virtual Reality) with the immersion and interactivity factors, the user interfaces (visiocasque, binaural...). In a second part it details the functions and techniques: captation, processing, coding and formatting, delivery and use. The last part specifies the coding and formatting standards for delivery, in particular those of the MPEG-I Immersive Media suite (OMAF...).

Mots-clés :

Vidéo 360°, audio 3D, Audio-visuel immersif, réalité, virtualité

Keywords:

360 video, 3D audio, immersive audio-visual, reality, virtuality

Plan détaillé

1	Principes et composantes - Définitions	5
1.1	Images	5
1.1.1	Vidéo 360°	5
1.1.2	Images infographiques CG	8
1.2	Sons	8
1.2.1	Formats génériques de représentation audio 3D	8
1.2.1.1	Canaux	9
1.2.1.2	Objets	9
1.2.1.3	Scènes	11
1.2.2	Binaural	13
1.3	Du réel au virtuel	15
1.3.1	Continuum de réalité-virtualité	15
1.3.2	Degrés de réalité-virtualité	15
1.3.2.1	Réalité virtuelle VR (<i>Virtual Reality</i>)	16
1.3.2.2	Réalité mixte MR (<i>Mixed Reality</i>)	16
1.3.2.3	Réalité augmentée AR (<i>Augmented Reality</i>)	16
1.3.2.4	Virtualité augmentée VA (<i>Augmented Virtuality</i>)	16
1.3.2.5	Réalité étendue XR (<i>eXtended Reality</i>)	16
1.3.2.6	Vidéo 360° // VR	17
1.3.3	Facteurs et degrés d'usage des systèmes audiovisuels interactifs	17
1.3.3.1	Immersion & présence	18
1.3.3.2	Interactivité	19
1.3.3.3	Degrés de liberté DoF (<i>Degrees of Freedom</i>)	19
1.4	Interfaces-utilisateurs et paramètres	21
1.4.1	Interfaces de visualisation	22
1.4.1.1	Visiocasques (HMD, <i>Head-Mounted Display</i>)	22
1.4.1.2	Fenêtre magique (<i>Magic Window</i>)	22
1.4.1.3	Second écran	23
1.4.1.4	Champ de vision (FOV, <i>Field of View</i>)	23
1.4.1.5	Fenêtre de visualisation (<i>viewport</i>)	25
1.4.2	Interfaces d'écoute	25
1.4.2.1	Haut-parleurs	26
1.4.2.2	Casques audio	26
2	Fonctions, systèmes et techniques	26
2.1	Workflow et fonctions	26
2.2	Systèmes vidéo 360°-I	28
2.2.1	Acquisition	29
2.2.2	Assemblage des images 360° multicaméra (<i>stitching</i>)	30
2.2.3	Projection Equirectangulaire ERP / Cubique CMP	32
2.2.4	Empaquetage en régions	33
2.2.5	Fisheye	34
2.2.6	Codage & formatage pour la distribution :	35
2.2.7	Encapsulation/désencapsulation fichier, ou segment DASH	36
2.2.8	Distribution (<i>delivery</i>)	36
2.2.9	Rendu (<i>rendering</i>)	36
2.3	Systèmes audio 3D	37
2.3.1	Systèmes de captation audio	38
2.3.1.1	Captation audio pour une représentation basée « canaux »	38
2.3.1.2	Captation audio pour une représentation basée « objets »	38
2.3.1.3	Captation audio pour une représentation basée « scènes »	38
2.3.2	Workflow de production de contenu audio 3D	40
2.3.2.1	Processus de production audio basée « scènes »	40
2.3.2.2	Processus de production audio basée « canaux »	41
2.3.2.3	Production et post-production audio basée « objets »	42

2.3.3	Formats de production audio spatiale en 360°-I & normes	43
2.3.3.1	Audio Definition Model (ADM)	43
2.3.3.2	Multi-Dimensional Audio (MDA)	43
2.3.3.3	AmbiX	44
2.3.3.4	AC-4.....	44
2.3.3.5	MPEG-H 3D Audio.....	44
2.3.3.6	Vers des métadonnées audio universelles ?	44
2.3.4	Système de rendu audio	45
2.3.4.1	Système de rendu audio basée « canaux ».....	45
2.3.4.2	Système de rendu audio basée « objets ».....	46
2.3.4.3	Système de rendu audio basée « scènes »	46
3	Normalisation des médias omnidirectionnels et immersifs	46
3.1	Formats de production	46
3.1.1	Formats de production vidéo 360°-I.....	47
3.1.2	Formats de production audio 3D.....	47
3.2	Formats de codage/compression (<i>coding</i>)	47
3.2.1	Vidéo	47
3.2.1.1	HEVC	47
3.2.1.2	VVC (Versatile Video Coding).....	47
3.2.2	Audio	47
3.2.3	Images fixes.....	47
3.3	MPEG-I : Représentation codée des médias immersifs	48
3.3.1	Une suite de normes	48
3.3.2	OMAF (<i>Omnidirectional Media Format</i>)	49
4	Conclusion	49

Introduction

Cet article propose un panorama des « médias immersifs » et des systèmes associés de production et de distribution. Le terme « immersif » est utilisé dans le cas des médias numériques pour décrire la capacité d'un système à absorber totalement un utilisateur dans une scène audiovisuelle. « Média immersif » peut être lié à du contenu naturel et/ou généré par ordinateur. Dans cet article (TE5690) et les deux suivants (TE5692 et TE5694), on s'intéresse exclusivement à la vidéo 360° et à l'audio 3D. L'immersion dans une scène est fortement renforcée par l'interactivité, qui peut aller d'un simple glissement de doigt sur une tablette pour découvrir une scène vidéo à 360° jusqu'à des rotations de la tête d'un utilisateur équipé d'un visiocasque pour naviguer dans la scène, accompagnées ou non de déplacement.

Depuis le premier procédé de projection d'images à 360°, sur un écran circulaire balayé par 10 projecteurs cinématographiques synchronisés disposés en cercle, fut le Cinéorama breveté en 1897 par Raoul Brimoin-Sanson... jusqu'au développement professionnel puis grand public de la vidéo 360° et des visiocasques à la fin du 20^e puis au début du 21^e siècle. Pour l'audio 3D, si on considère la stéréo comme un ancêtre, on peut rappeler la retransmission sur lignes téléphoniques d'un opéra par Clément Ader en 1881, avec le Théâtrophone... jusqu'à l'émergence des effets audio en 3D dans les années 1990, dans les PC et les consoles de jeux. Quant à l'un des premiers systèmes immersifs et interactifs, on peut citer le Sensorama, breveté en 1962 par Morton Heilig, qui était un simulateur de balade en moto à travers New York, avec un écran, des générateurs de sons, de vent, d'odeurs... jusqu'aux des plus récents des simulateurs de pilotage d'avion ou le simulateur de « vol d'oiseau » immersif Birdly (2016).

Le développement des médias immersifs a été favorisé par l'essor du numérique, en particulier dans le domaine des formats de compression de la vidéo numérique et des réseaux de distribution Internet, ainsi que par la diffusion des jeux vidéo et des dispositifs portables (visiocasques, tablettes et smartphones).

Parmi les applications des médias immersifs 360°, on peut citer : la diffusion d'événements (en parallèle ou non à un flux télévision), la cinématographie linéaire ou interactive, des applications éducatives, la communication interpersonnelle et de groupe (par exemple vidéoconférence, ou avec les jeux en ligne, ou dans un lieu de rencontre virtuel), les visites guidées, le visionnage de contenu à la demande sur

visiocasque...

Au-delà de l'intérêt lors de la découverte d'une nouvelle technologie, la vidéo 360° et l'audio 3D créent une sensation de présence dans un espace (dans une autre dimension), de liberté (de choix du point de vue), d'implication dans une histoire. Toutefois, les médias immersifs sont sujets à des contraintes : la nécessité de normes et de plateformes de distribution, la capacité des réseaux pour transmettre le volume important de données, la perte de repères de l'utilisateur et le malaise sensoriel qui éventuellement en découle...

Malgré un regain d'intérêt en particulier dans les années 2010, les contraintes techniques, économiques et d'usage n'ont pas permis la progression espérée. La série des trois articles propose donc un panorama des éléments techniques au niveau des médias numériques vidéo et audio, au niveau des systèmes et des processus de production et de distribution, pouvant rendre possible un nouvel essor. La série des 3 articles propose donc un panorama des éléments techniques relatifs aux médias immersifs vidéo et audio, ainsi qu'aux systèmes et processus de production et de distribution, pouvant rendre possible un nouvel essor.

Cet article est le premier de la série de trois. Il est composé de trois parties :

- *la première partie définit d'abord les composantes des systèmes audio-visuels immersifs : le contenu omnidirectionnel (vidéo à 360° et audio 3D), le continuum du réel au virtuel et ses divers degrés : AR (Augmented Reality), MR (Mixed Reality), AV (Augmented Virtuality), VR (Virtual Reality), avec les facteurs d'immersion et d'interactivité, les interfaces-utilisateurs (visiocasque, binaural...);*
- *dans la deuxième partie, il détaille les fonctions et les techniques : l'acquisition, le traitement, le codage et formatage, la distribution et l'utilisation ;*
- *la dernière partie précise les normes de codage et de formatage, en particulier celles de l'OMAF (Omnidirectional MediA Format) dans l'ensemble MPEG-I (Immersive Media).*

La lecture de ce premier article est indispensable pour la bonne compréhension des deux articles suivants. De nombreux renvois entre les 3 articles sont insérés dans chacun des articles.

Le second article, TE5692, détaille les étapes du processus de production et de post-production ainsi que les outils associés. Il propose ensuite quelques recommandations et bonnes pratiques de production de vidéo 360° et d'audio 3D, ainsi que des exemples de formats de production.

Le 3^e article, TE5694, décrit les formats, profils et normes utilisés pour la distribution en fichier et/ou sur réseau, illustrés par deux études de cas. Il analyse ensuite les aspects liés à la « consommation » de produits audiovisuels immersifs, tels que les interfaces et le malaise sensoriel. Après un panorama d'applications et de cas d'usage, il évalue les facteurs de qualité vidéo et audio et ceux de la qualité d'expérience.

Nota 1 : Le lecteur trouvera en fin d'article une liste des principaux acronymes et abréviations utilisés tout au long de ce document. De nombreux termes et le texte de quelques figures ont été conservés en anglais (en caractères italiques), soit parce qu'ils sont couramment utilisés dans le milieu professionnel, soit pour éviter toute ambiguïté de la traduction française qui en est proposée, soit enfin pour aider à la lecture des spécifications ainsi que d'autres documents en anglais.

Nota 2 : Pour le lecteur peu initié à la vidéo à 360° et à l'audio 3D, des sites Web sont référencés en fin de la partie annexe « Pour en savoir plus », où il pourra en visionner et en écouter.

TE5692

Version préliminaire - Ne pas diffuser

Médias immersifs : vidéo 360° & audio 3D
- Production : processus, outils, recommandations

*Immersive media: 360 video & 3D audio -
Production: workflow, tools, guidelines*

Jean-Noël GOUYET

Ingénieur-formateur et consultant en techniques et gestion des médias numériques.
Ancien chargé d'études à la Direction de la Recherche de l'INA.

Résumé

Cet article est le second d'une série de trois, et concerne la production pour les systèmes audio-visuels immersifs. Il décrit d'abord le *workflow*, le processus de production et de post-production des médias incluant la vidéo à 360° et l'audio 3D, Puis il détaille les outils associés à la captation (caméras 360°, microphones) puis au traitement de ces médias immersifs. Il propose ensuite des recommandations et bonnes pratiques pour l'acquisition et la création vidéo et audio. La dernière partie spécifie des formats de production vidéo et audio.

Abstract

This article is the second of a series of three and concerns the production for immersive audio-visual systems. It firstly describes the workflow, i.e. the production and post-production process for media including 360° video and 3D audio. It secondly details the tools associated with the capture (360° cameras, microphone arrays) and processing of these immersive media. It then proposes guidelines and best practices for video and audio acquisition and creation. The last part specifies video and audio production formats.

Mots-clés :

Vidéo 360°, audio 3D, audio-visuel immersif ; production

Keywords:

360 video, 3D audio, immersive audio-visual; production

Plan détaillé

1	Processus de production & postproduction	6
1.1	Workflow	6
1.2	Etapes	7
1.2.1	Pré-production	7
1.2.1.1	Ecriture et <i>storyboard</i> pour tournage de vidéo 360°	7
1.2.1.2	Liste des plans tournés et plan de tournage	9
1.2.2	Production	9
1.2.2.1	Réalisation, mise en scène 360°	9
1.2.2.2	Visionnage de contrôle en direct sur le lieu de tournage	10
1.2.2.3	Installation de la scène	10
1.2.2.4	Rapport script	10
1.2.3	Postproduction	10
1.2.3.1	Assemblage sommaire des images multicaméra 360°	11
1.2.3.2	Montage	11
1.2.3.3	Assemblage final des images multicaméra 360°	11
1.2.3.4	Effets visuels, graphismes et titres	12
1.2.3.5	Conception sonore	12
1.2.3.6	Rendu final et distribution	12
2	Outils pour la production et la post-production	13
2.1	Outils d'acquisition	13
2.1.1	Caméras 360°	13
2.1.1.1	Plateforme caméra panoptique	13
2.1.1.2	Plateforme de caméras à miroir	14
2.1.1.3	Caméras « <i>fish-eye</i> »	15
2.1.1.4	Caméras <i>light-field</i>	16
2.1.2	Microphones	17
2.1.2.1	Microphones basées « canaux »	17
2.1.2.2	Enregistrement binaural	19
2.1.2.3	Ambisonie	20
2.1.2.4	Enregistrement basé objets	23
2.2	Outils logiciels	23
2.2.1	<i>Stitching</i>	23
2.2.2	Post-production	26
2.2.2.1	Fonctions et fonctionnalités en post-production	26
2.2.2.2	Outils de post-production vidéo 360°-I	27
2.2.2.3	Outils de post-production audio 3	27
3	Recommandations pour la production de vidéo 360° et d'audio 3D	28
3.1	Expérience utilisateur et défis de production	28
3.2	Recommandations pour l'acquisition & la création vidéo	29
3.2.1	Champ de vision	29
3.2.2	Position du contenu	29
3.2.3	Mouvements de caméra	30
3.2.4	Fréquence-image de captation et fidélité de captation du mouvement	31
3.2.5	Orientation	31
3.2.6	Hauteur perçue de l'oeil du spectateur	31
3.2.7	Proximité des objets dans la scène	31
3.2.8	Contenu 3D stéréoscopique	32
3.2.9	Contenu HDR	32
3.2.10	Durée du contenu	33
3.2.11	Sous-titres	33
3.2.11.1	Position	33

	3.2.11.2 Orientations de production	34
3.3	Recommandations pour l'acquisition & le traitement audio	35
3.3.1	Captation	35
3.3.2	Ambisonie.....	35
3.3.3	Traitement.....	36
3.3.4	Audio diégétique / non diégétique.....	36
3.3.5	Niveau sonore.....	36
3.3.6	Présentation / rendu pour le mixage et le contrôle (<i>monitoring</i>).....	36
4	Formats de production Vidéo 360° / Audio 3D & paramètres	37
4.1	Formats <i>Master</i>	37
4.1.1	Format <i>Master</i> Vidéo.....	37
4.1.1.1	Définition d'image.....	37
4.1.1.2	Projection.....	37
4.1.1.3	Rapport géométrique de format d'image – Couverture (<i>Coverage</i>).....	38
4.1.1.4	Fréquences-images images/seconde.....	38
4.1.1.5	Métadonnées vidéo.....	38
4.1.1.6	Formats de fichier <i>Master</i>	39
4.1.2	Format <i>Master</i> Audio.....	40
4.1.2.1	Métadonnées audio	40
4.1.2.2	Format audio	40
4.1.2.3	Profondeur de codage et fréquence d'échantillonnage.....	40
4.2	Format de production et d'échange international en <i>broadcasting</i>	40
4.2.1	Caractéristiques des images 360° dans les applications 3DoF.....	41
4.2.2	Projection de vidéo omnidirectionnelle.....	41
4.2.2.1	Structure de projection et système de coordonnées	42
4.2.2.2	Formats de projection omnidirectionnelle	42
4.2.3	Caractéristiques de présentation d'images 360°	43
5	Conclusion	43

TE5694

Version préliminaire - Ne pas diffuser

Médias immersifs : vidéo 360° & audio 3D

- Distribution. Usages. Qualité

Immersive media: 360 video & 3D audio

- Delivery. Use cases. Quality

Jean-Noël GOUYET

Ingénieur-formateur et consultant en techniques et gestion des médias numériques.
Ancien chargé d'études à la Direction de la Recherche de l'INA.

Résumé

Cet article est le troisième et dernier article d'une série de trois, concernant les systèmes audio-visuels immersifs. Il décrit d'abord les formats et les normes nécessaires au codage et au formatage des médias immersifs (vidéo 360°, audio 3D), en vue de leur distribution. Deux cas d'étude illustrent d'une part le *streaming* ou le téléchargement de contenu 360°, et d'autre part l'association contenu 360° et télévision en direct. Puis, l'article présente un panorama d'applications et cas d'usage et décrit les interfaces-utilisateurs nécessaires à l'utilisation des expériences audio-visuelles immersives. Il explicite enfin les facteurs de la qualité vidéo et audio, ainsi que ceux de la qualité d'expérience.

Abstract

This article is the third and last one of a series of three about immersive audio-visual systems. It first describes the formats and standards needed to encode and format immersive media (360° video, 3D audio) for distribution. Two case studies illustrate on the one hand the streaming or downloading of 360° content, and on the other hand the association of 360° content and live TV. Then, the article presents a panorama of applications and use cases and describes the user interfaces required to "consume" immersive audio-visual experiences. Finally, it explains the factors of video and audio quality, as well as those of the quality of experience.

Mots-clés :

Vidéo 360°, audio 3D, audio-visuel immersif, médias immersifs ; distribution, usages, qualité

Keywords:

360 video, 3D audio, immersive audio-visual, immersive media; delivery, use cases, quality

Plan détaillé

1	Distribution	6
1.1	Compression des médias immersifs.....	7
1.1.1	Formats de compression vidéo.....	7
1.1.1.1	H.265 / HEVC (<i>High-Efficiency Video Coding</i>).....	7
1.1.1.2	H.266 / VVC (<i>Versatile Video Coding</i>).....	8
1.1.1.3	Compression à nuage de points.....	8
1.1.2	Formats de compression audio	8
1.2	Formats de distribution & normalisation	8
1.2.1	OMAF (<i>Omnidirectional Media Format</i>)	9
1.2.2	DASH (<i>Dynamic Adaptive Streaming over HTTP</i>).....	11
1.2.3	MMT (<i>MPEG Media Transport</i>)	13
1.2.4	MBMS (<i>Multimedia Broadcast Multicast Services</i>)	13
1.2.5	Organismes de normalisation et de recommandations.....	13
1.2.5.1	3GPP - ETSI.....	14
1.2.5.2	VR Industry Forum.....	14
1.2.5.3	ITU	14
1.3	Profils média	14
1.3.1.1	Profils Vidéo	14
1.3.1.2	Profils audio.....	17
1.3.1.3	Profils d'image fixe.....	18
1.3.1.4	Profils OMAF de texte synchronisé	18
1.4	Profils de présentation	19
1.5	Etude de cas 1 : <i>Streaming</i> ou téléchargement de contenu 360°-I.....	19
1.5.1	Description.....	19
1.5.2	Architecture de référence	20
1.5.3	Solutions techniques	21
1.5.3.1	Formats et profils appropriés.....	21
1.5.3.2	Systèmes de distribution.....	22
1.6	Etude de cas 2 : Télévision + 360°-I en direct	23
1.6.1	Description.....	23
1.6.2	Architecture de référence et workflow	23
1.6.3	Solutions techniques	25
1.6.3.1	Système de production pour des événements en direct.....	25
1.6.3.2	Profils média et de présentation appropriés.....	25
1.7	Diffusion d'audio-visuel immersif AIAV	25
2	Applications et usages.....	26
2.1	Points d'observation.....	26
2.1.1	Point d'observation unique	26
2.1.2	Point d'observation variable en continu	26
2.2	Evènement distribué en <i>broadcast/multicast</i>	27
2.3	<i>Streaming</i> 360°-I à la demande.....	28
2.4	Distribution d'une collection de contenus 360°-I	28
2.5	Services de direct (<i>live</i>) sur visiocasque	28
2.6	TV "sociale" et 360°-I	28
2.7	Cas d'usage de 360°-I cinématographique.....	28
2.8	Cas d'usage d'applications éducatives.....	29
2.9	Cas d'usage avec appels visio 360°-I.....	30
2.10	Cas d'usage avec du contenu généré par l'utilisateur	30
2.11	Communiquer dans un monde virtuel.....	31
2.12	Utilisation de contenu TV sur visiocasque	32
2.13	Cas d'utilisation de régions d'intérêt dans la vidéo 360°-I.....	33
2.14	Expériences et productions de 360°-I en télévision	34

3	Utilisation	35
3.1	Défis pour l'utilisation	35
3.1.1	Qualité	35
3.1.2	Malaise sensoriel (<i>sensory sickness</i>).....	36
3.1.3	Présentation du contenu	36
3.1.4	Plateformes de distribution 360°-I.....	37
3.2	Interfaces-utilisateurs	37
3.2.1	Interfaces de visualisation	37
3.2.1.1	Visiocasques VR (HMD, <i>Head-Mounted Display</i>)	37
3.2.1.2	Lunettes AR.....	39
3.2.2	Interfaces de contrôle.....	41
3.2.3	Interfaces-utilisateurs d'écoute	41
3.2.3.1	Casques audio.....	41
3.2.3.2	Ecouteurs intra-auriculaires	42
3.2.3.3	Optimisation	43
3.2.3.4	Systèmes de haut-parleurs externes	43
3.2.3.5	Bluetooth	43
4	Qualité	43
4.1	Evaluation de la qualité vidéo.....	43
4.1.1	Facteurs de qualité visuelle et mesures	43
4.1.1.1	Qualité du contenu vidéo à la restitution.....	44
4.1.1.2	Erreurs de « <i>stitching</i> »	45
4.1.1.3	Qualité de l'immersion visuelle	45
4.1.2	Tests.....	46
4.2	Evaluation de la qualité audio	47
4.2.1	Facteurs de qualité audio	47
4.2.2	Tests	47
4.2.2.1	Evaluation de la qualité d'audio basée « scènes »	47
4.2.2.2	Evaluation de la qualité d'audio basée « objets »	48
4.2.2.3	Evaluation de la qualité d'audio basée « canaux »	48
4.3	Aspects de latence et de synchronisation	48
4.3.1	Latence d'interaction	48
4.3.1.1	Interaction vidéo	48
4.3.1.2	Interaction audio	49
4.3.2	Synchronisation audio/vidéo	49
4.4	Qualité d'expérience (QoE).....	49
4.4.1	Facteurs de contenu	49
4.4.1.1	Impact du contenu vidéo sur la qualité d'expérience.....	49
4.4.1.2	Qualité de l'expérience audio	50
4.4.2	Facteurs réseau et terminal.....	51
4.4.2.1	Impact du réseau sur la qualité d'expérience.....	51
4.4.2.2	Impact de l'équipement utilisateur	51
5	Conclusion	52