Supplementary Material:

Exploring the impact of 360° movie cuts in users' attention

Carlos Marañes, Diego Gutierrez, Ana Serrano

Universidad de Zaragoza

This document offers additional information and details in the following topics of the main paper:

- Types of cuts
- ROI configuration
- Analyzing the influence of cuts
- Analyzing the influence of cuts without gestural and auditory cues.
- Analyzing the influence of the used system.

A Types of cuts

We have manually labeled our cuts, helping us with the IOC and AUC curves. In Table S1 we show each cut labeled and the corresponding time stamp.

B ROI configuration

We have defined the region of the scene that contains a ROI with a box. As the ROI is motionless or moves virtually nothing it always remains centered in the box. Fig. S1 illustrates the area the area delimiting the ROI.

C Analyzing the influence of cuts

We display in Fig. S2 the Inter-Observer Congruency (IOC) computed as described in the main paper for all analyzed cuts.

D Analyzing the influence of cuts without gestural and auditory cues.

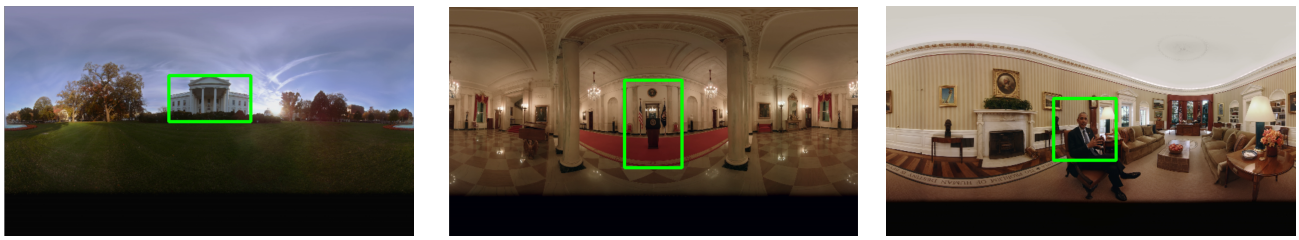
Additionally, we have repeated the same statistical analysis shown in the main paper for cuts in which there is no gestural and auditory cues involved. We have selected 11 cuts being distributed as follows: $ROI \rightarrow ROI$: three cuts, $nROI \rightarrow ROI$: three cuts, $nROI \rightarrow nROI$: three cuts, $ROI \rightarrow nROI$: two cuts. We display in Fig. S3 the average of each metric per different types of cuts. Fig. S4 shows the averages per each type of cut for the metrics *nFix*, *traveledDistance* and *percSceneWatched*. Finally, Tables S2 and S3 show the significance tests by using a Generalized Linear Mixed Model as the main paper explains. Results follow the same tendency as the ones presented in the main paper.

E Analyzing the influence of the used system

Even though it is out of the scope of this work, we present the results obtained also adding in the statistical analysis described in the main paper the system used by the user as a predictor ($system = \{Computer, Mobile\}$). There are 967 users who meet $system = Mobile$ and 2156 users who meet $system = Computer$. There are 136 whose system is unknown and have not been included in the analysis. Results are shown in Table S4 and in Fig. S5. Results follow the same tendency as the ones presented in the main paper. However, we had no control about how the system has been used.

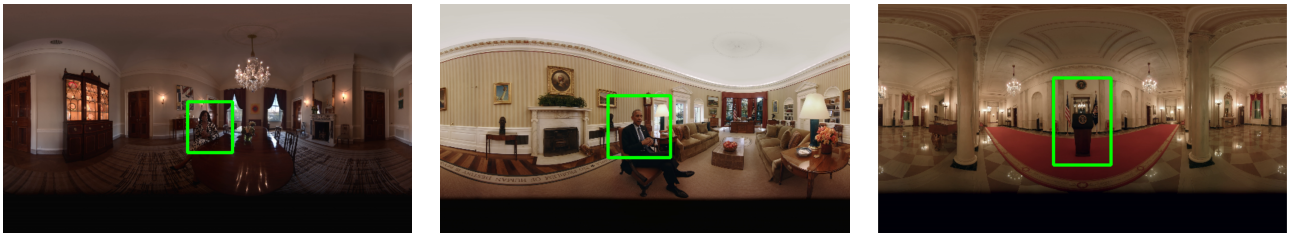
Cut name	S_b	S_a	**Brief description**	**Starting time stamp**	**Ending time stamp**
001-north	ROI	ROI	The White House	00:17	00:29
003-podium	ROI	ROI	Speaker's podium in a corridor	01:16	01:28
005-obama	ROI	ROI	Obama talking	02:02	02:16
021-michelle	ROI	ROI	Michelle talking	08:40	08:51
047-podium	ROI	ROI	Speaker's podium in a corridor	18:38	18:51
015-obama	nROI	ROI	Obama talking	06:50	07:02
019-rosegarden	nROI	ROI	Person playing with dogs in a garden	08:06	08:19
035-michelle	nROI	ROI	Michelle talking	13:39	13:52
045-obama	nROI	ROI	Obama talking	17:51	18:04
009-situation	nROI	nROI	Room	04:17	04:30
011-roosevelt	nROI	nROI	Room	05:06	05:19
025-green	nROI	nROI	Room	10:07	10:20
027-east	nROI	nROI	Lounge	10:44	10:57
029-vermeil	nROI	nROI	Room	11:27	11:40
007-cabinet	ROI	nROI	Cabinet	03:10	03:23
017-oval	ROI	nROI	Room	07:18	07:31
023-red	ROI	nROI	Room	09:30	09:43
037-blue	ROI	nROI	Room	14:32	14:45

Table S1: Each cut labeled and a brief description of the scene after the cut. Starting time stamp indicates the minute and second in the video where the cut starts and ending time stamp where the cut ends. For the analyses, only the first 6 seconds of the scene after the cut (S_b) has been analyzed.



(a) 001-north **(b)** 003-podium **(c)** 005-obama

(d) 015-obama **(e)** 019-rosegarden **(f)** 021-michelle



(g) 035-michelle **(h)** 045-obama **(i)** 047-podium

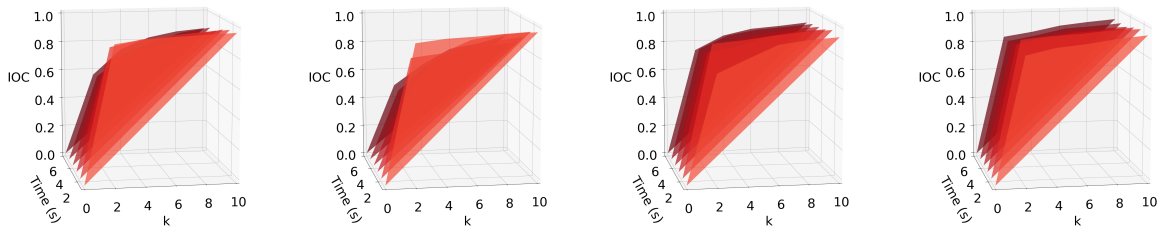
Figure S1: ROIs of the scenes delimeted with boxes which belongs to cuts which are $S_a = ROI$.

Metric	Source	F	df1	df2	Sig.
framesToROI	S_b	600,257	1	16293	,000
percFixInside	S_b	170,166	1	15798	,000
nFix	S_b	125,386	1	35845	,000
	S_a	4390,463	1	35845	,000
	$S_b * S_a$	154,501	1	35845	,000
traveledDistance	S_b	220,682	1	35845	,000
	S_a	2812,226	1	35845	,000
	$S_b * S_a$	486,511	1	35845	,000
percSceneWatched	S_b	78,427	1	35845	,000
	S_a	4228,962	1	35845	,000
	$S_b * S_a$	42,624	1	35845	,000

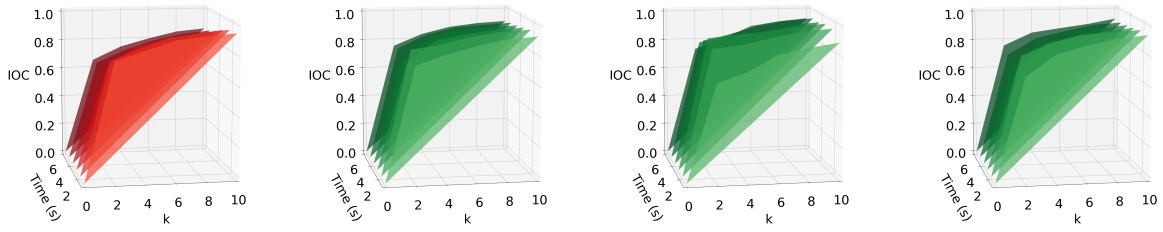
Table S2: Tests of fixed effects for the different metrics for those cuts which do not involve gestural and auditory cues.

Metric	S_a	S_b	t	df	Adj. Sig
nFix	ROI	ROI - nROI	-17,467	35845	,000
		nROI - ROI	17,467	35845	,000
	nROI	ROI - nROI	,790	35845	,430
		nROI - ROI	,790	35845	,430
	ROI - nROI	ROI	47,806	35845	,000
	nROI - ROI		-47,806	35845	,000
	ROI - nROI	nROI	66,969	35845	,000
	nROI - ROI		-66,969	35845	,000
traveledDistance	ROI	ROI - nROI	21,246	35845	,000
		nROI - ROI	-21,246	35845	,000
	nROI	ROI - nROI	-7,529	35845	5.240E-14
		nROI - ROI	7,529	35845	5.240E-14
	ROI - nROI	ROI	-21,610	35845	,000
	nROI - ROI		21,610	35845	,000
	ROI - nROI	nROI	-49,408	35845	,000
	nROI - ROI		49,408	35845	,000
percSceneWatched	ROI	ROI - nROI	9,404	35845	,000
		nROI - ROI	-9,404	35845	,000
	nROI	ROI - nROI	1,558	35845	,119
		nROI - ROI	-1,558	35845	,119
	ROI - nROI	ROI	-39,923	35845	,000
	nROI - ROI		39,923	35845	,000
	ROI - nROI	nROI	-55,447	35845	,000
	nROI - ROI		55,447	35845	,000

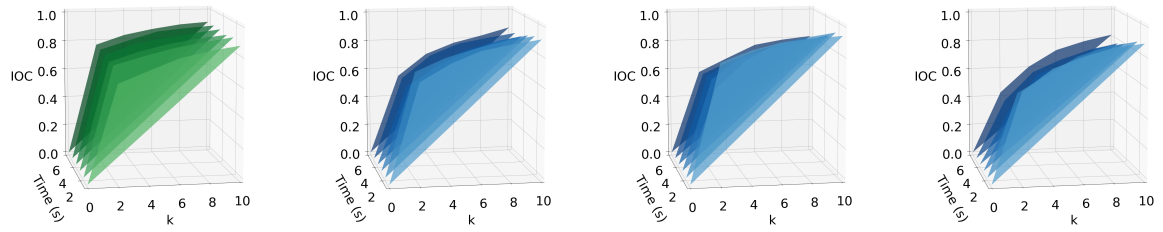
Table S3: Significance of pairwise comparisons for the set of cuts which do not include gestural and auditory cues computed by using a Generalized Linear Mixed Model for the all of the metrics for the interaction $S_a * S_b$.



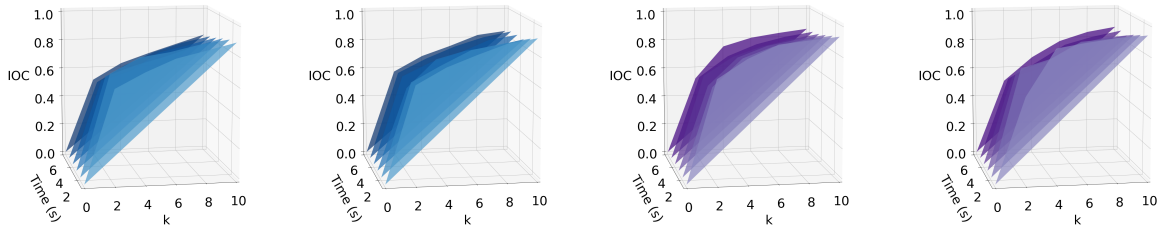
(a) 001-north **(b)** 003-podium **(c)** 005-obama **(d)** 021-michelle



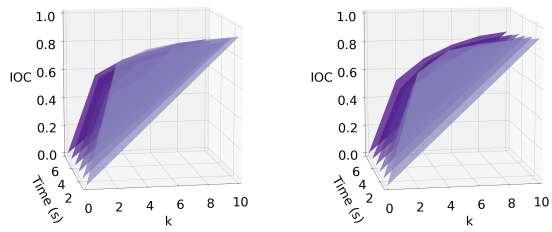
(e) 047-podium **(f)** 015-obama **(g)** 019-rosegarden **(h)** 035-michelle



(i) 045-obama **(j)** 009-situation **(k)** 011-roosevelt **(l)** 025-green



(m) 027-east **(n)** 029-vermeil **(o)** 007-cabinet **(p)** 017-oval



(q) 023-red **(r)** 037-blue

Figure S2: IOC for each cut. **(a), (b), (c), (d), (e)**: $ROI \rightarrow ROI$. **(f), (g), (h), (i)**: $nROI \rightarrow ROI$. **(j), (k), (l), (m), (n)**: $nROI \rightarrow nROI$. **(o), (p), (q), (r)**: $ROI \rightarrow nROI$.

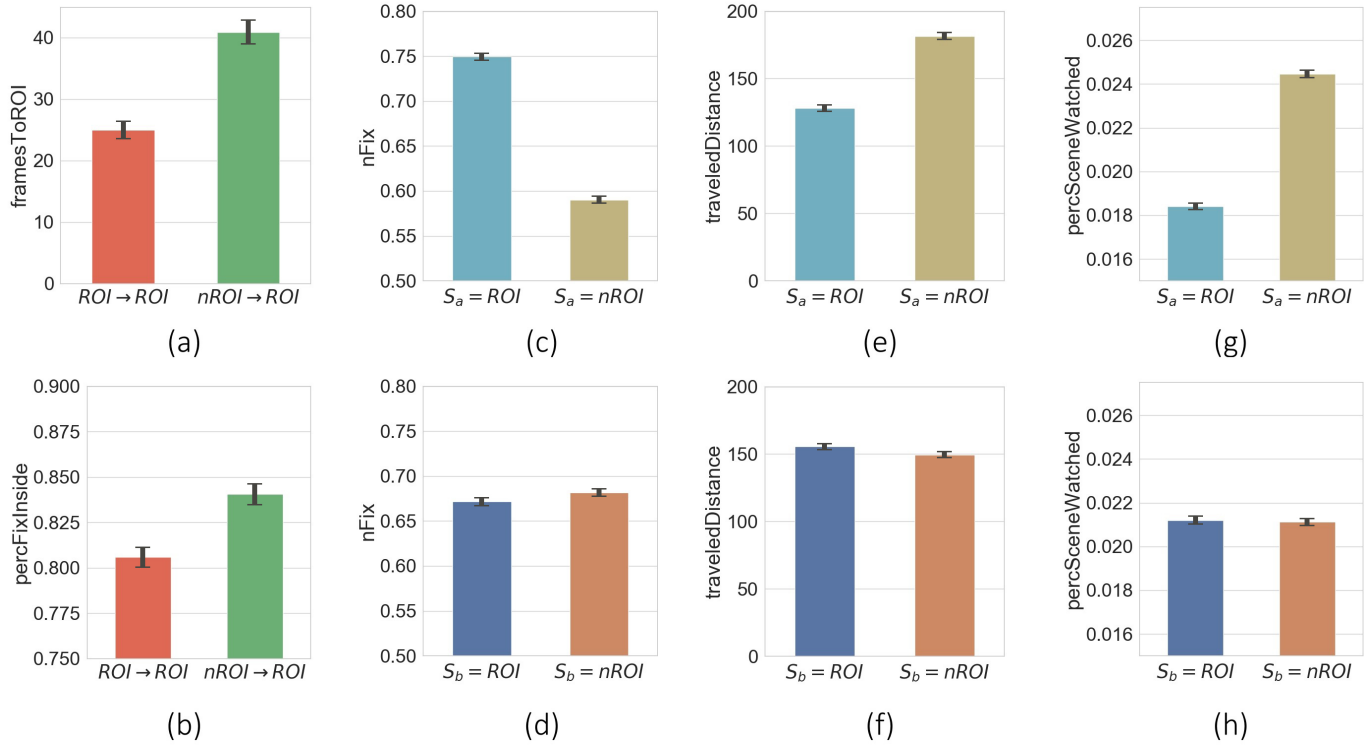


Figure S3: **(a)**: Mean *framesToROI* for the cuts $\{ROI, nROI\} \rightarrow ROI$. **(b)**: Mean *percFixInside* for the cuts $\{ROI, nROI\} \rightarrow ROI$. **(c)**: Mean *nFix* gruped by type of scene after the cut (S_a). **(d)**: Mean *nFix* grouped by type of scene before the cut (S_b). **(e)**: Mean *traveledDistance* grouped by type of scene after the cut (S_a). **(f)**: Mean *traveledDistance* grouped by type of scene before the cut (S_b). **(g)**: Mean *percSceneWatched* grouped by type of scene after the cut (S_a). **(h)**: Mean *percSceneWatched* grouped by type of scene before the cut (S_b). Error bars correspond to a 95% confidence interval. Analyzed cuts do not include gestural and auditory cues.

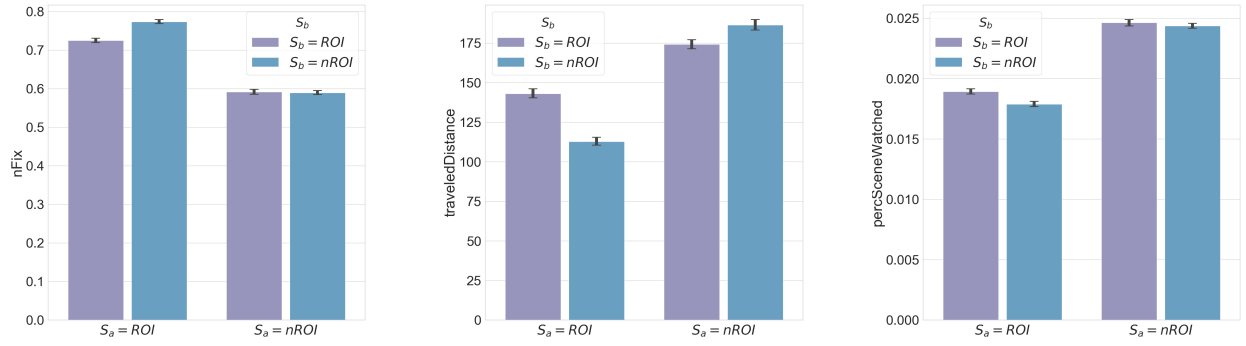


Figure S4: **(a)**: Mean *nFix* metric for each type of cut. **(b)**: Mean *traveledDistance* metric for each type of cut. **(c)**: Mean *percSceneWatched* for each type of cut. Error bars correspond to a 95% confidence interval. Analyzed cuts do not include gestural and auditory cues.

Metric	Source	F	df1	df2	Sig.
framesToROI	S_b	257,719	1	28103	,000
	$system$	1,476	1	28103	,224
	$S_b * system$	12,974	1	28103	,000
percFixInside	S_b	720,842	1	27255	,000
	$system$	223,849	1	27255	,000
	$S_b * system$	161,957	1	27255	,000
nFix	S_b	444,841	1	56207	,000
	S_a	5832,371	1	56207	,000
	$system$	344,334	1	56207	,000
	$S_b * S_a$	29,547	1	56207	,000
	$S_b * system$	129,722	1	56207	,000
	$S_a * system$	187,688	1	56207	,000
traveledDistance	S_b	385,402	1	56207	,000
	S_a	3567,424	1	56207	,000
	$system$	237,647	1	56207	,000
	$S_b * S_a$	238,609	1	56207	,000
	$S_b * system$	53,865	1	56207	,000
	$S_a * system$	4,031	1	56207	,045
percSceneWatched	S_b	54,093	1	56207	,000
	S_a	5006,533	1	56207	,000
	$system$	126,403	1	56207	,000
	$S_b * S_a$	135,418	1	56207	,000
	$S_b * system$	74,219	1	56207	,000
	$S_a * system$	7,337	1	56207	,007

Table S4: Tests of fixed effects for the different metrics.

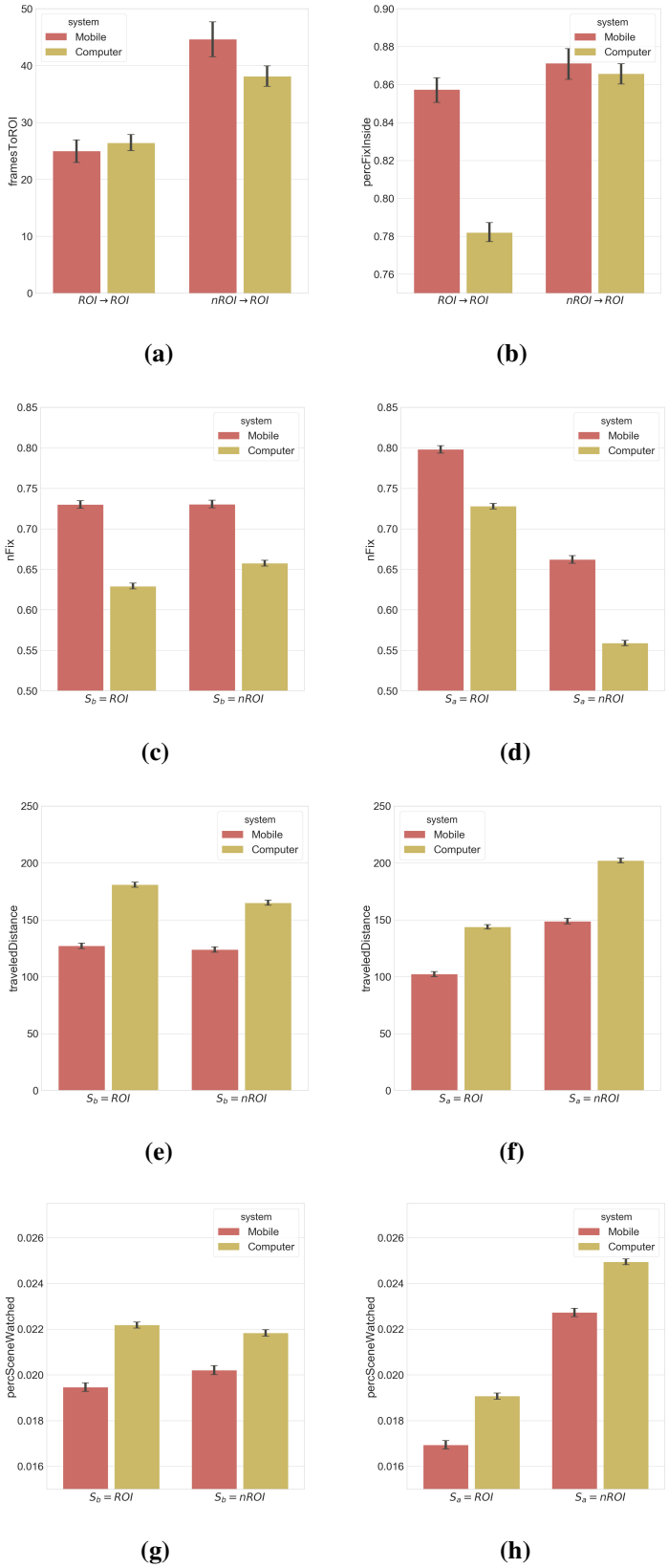


Figure S5: **(a)**: Mean *framesToROI* for the cuts $\{ROI, nROI\} \rightarrow ROI$ and per system. **(b)**: Mean *percFixInside* for the cuts $\{ROI, nROI\} \rightarrow ROI$ and per system. **(c)**: Mean *nFix* grouped by type of scene before the cut (S_b) and per system. **(d)**: Mean *nFix* gruped by type of scene after the cut (S_a) and per system. **(e)**: Mean *traveledDistance* grouped by type of scene before the cut (S_b) and per system. **(f)**: Mean *traveledDistance* grouped by type of scene after the cut (S_a) and per system. **(g)**: Mean *percSceneWatched* grouped by type of scene before the cut (S_b) and per system. **(h)**: Mean *percSceneWatched* grouped by type of scene after the cut (S_a) and per system. Error bars correspond to a 95% confidence interval.