variable name ScoreMemTest	Definition Count of correct answer (0 - 6)	How's record organised Per participant	What is better in terms of mean Higher	f Test GLMM (poisson)	Mean_mono 4.345238	Mean_binaural	Mean_head 4.416667	main_effect (sound) No	main_effect (sound, LMM) No	Mean_no_mask 4.222222	SD_no_mask 1.238637	Mean_mask 4.642857	SD_mask 1.238637	main_effect (mask)	main_effect (mask, LMM) F(2, 205) = 8.806 p = 0.003	(Loic) A Wilcoxon signed-rank tests was performed to compare the difference of score to the memory test in mask and no mask visual conditions. There was a significant difference in the score for mask (M=4.64, SD=1.39) and no mask (M=4.22, SD=1.24) conditions with Z=-2.698, p = .007, and a small effect	plot		interaction_effect (mask) & sound, LMM) No
MemConfidence	Likert's scale (1 - 5)	Per participant	Higher	GLMM (ordinal)	3.392857	3.071429	3.380952	No	No	3.293651	1.117622	3.269841	1.117622	No	No	size (r=-0.24). These results indicate that when offline conferees wore surgical masks, participants performed better at the memory test (M=4.64) than did participants in the conditions where offline conferees wore no mask (M=4.22).	The control of the co	No No	No
ScoreCompTest	Count of correct answer (0 - 6)	Per participant	Higher	GLMM (poisson)	2.345238	2.488095	2.857143	No	F(2, 205) = 18.006 p < 0.001	2.619048	0.577680	2.507937	0.577680	No	No	(Loic) [without mask]	Sound Sound	No	F(2, 205) = 11.423 p < 0.001
																A non-parametric Friedman's test of differences among repeated measures was conducted to compare the score to the comprehension test in the mono (M=2.26, SD=.627), spatial (M=2.79, SD=.47) and spatial with automatic head rotation (M=2.81, SD=.455) auditory conditions, and in the visual condition where the conferees did not wear a mask. The test rendered a χ^2 (2) value of 23.128 which was significant with p<.001. Post hoc analysis with Wilcoxon signed-rank tests was then performed. It resulted that there was a significant difference between mono and spatial audio (Z=-3.554, p<.001) with a large effect size (r=-0.548), and between mono and spatial audio with automatic head rotation (Z=-3,518, p<.001) with a large effect size (r=-0.543). These results indicate that in the condition where offline conferees did not wear surgical masks and with spatial audio with or without automatic head rotation, participants performed better at the comprehension test than with mono audio. [with mask] A non-parametric Friedman's test of differences among repeated measures was conducted to compare the score to the comprehension test in the mono (M=2.43, SD=.703), spatial (M=2.19, SD=.773) and spatial with automatic head rotation (M=2.9, SD=.370) auditory conditions, and in the visual condition where the conferees wore a mask. The test rendered a χ^2 (2) value of 23.009 which was significant with p<.001. Post hoc analysis with Wilcoxon signed-rank tests was then performed. It resulted that there was a significant difference between mono and spatial audio with head rotation (Z=-3.123, p=.002) with a moderate effect size (r=-0.482), and between spatial audio without automatic head rotation and spatial audio with automatic head rotation and spatial audio with automatic head rotation, participants performed better at the comprehension test than with mono or spatial audio without automatic head rotation, participants performed better at the comprehension test than with mono or spatial audio without automatic head rotation.	136- 136- 136- 136- 136- 136- 136- 136-		
CompConfidence	Likert's scale (1 - 5)	Per participant	Higher	GLMM (ordinal)	3.642857	3.297619	4.190476	No	F(2,205) = 15.462 p < 0.001	3.817460	1.022943	3.603175	1.022943	No	No	(Loic) [without mask] A non-parametric Friedman's test of differences among repeated measures was conducted to compare the comprehension confidence in the mono, SD=.909), spatial (M=3.6, SD=1.211) and spatial with automatic head rotation (M=4.24, SD=.790) auditory conditions, and in the visual condition where the conferees did not wear a mask. The test rendered a χ^2 (2) value of 10.606 which was significant with p=.005. Post hoc analysis with Wilcoxon signed-rank tests was then performed. It resulted that there was a significant difference between mono and spatial audio with automatic head rotation (Z=-2.925, p=.003) with a moderate effect size (r=-0.451), and between spatial audio without automatic head rotation and spatial audio with automatic head rotation and spatial audio with automatic head rotation (Z=-3.028, p=.002) with a moderate effect size (r=-0.467). These results indicate that in the condition where offline conferees did not wear surgical masks and with spatial audio with automatic head rotation, participants had a higher confidence in their answers to the comprehension test than with mono or spatial audio with automatic head rotation. [with mask] A non-parametric Friedman's test of differences among repeated measures was conducted to compare the comprehension confidence in the mono (M=3.67, SD=1.097), spatial (M=3, SD=1.230) and spatial with automatic head rotation (M=4.14, SD=1.260) auditory conditions, and in the visual condition where the conferees wore a mask. The test rendered a χ^2 (2) value of 17.671 which was significant with p<0.001. Post hoc analysis with Wilcoxon signed-rank tests was then performed. It resulted that there was a significant difference between mono and spatial audio (Z=-2.647, p=.008) with a moderate effect size (r=-0.408), and between spatial audio without automatic head rotation and spatial audio with automatic head rotation led to higher confidence in the participant answer to the comprehension test than spatial audio without automatic head rotation led to hi		χ2(2, N = 252)= ,p = 0.03573	No
PostSpeakerIdent	Likert's scale (1 - 5)	Per participant	Higher	GLMM (ordinal)	3.702381	4.047619	3.511905	No	F(2,205) = 8.745 p < 0.001	4.238095	0.804274	3.269841	0.804274	No	F(2,205) = 83.385 p < 0.001	(Loic) [without mask] A non-parametric Friedman's test of differences among repeated measures was conducted to compare the difficulty to identify the speaking conferee in the mono (M=4.26, SD=.734), spatial (M=4.45, SD=.670) and spatial with automatic head rotation (M=4, SD=.937) auditory conditions, and in the visual condition where the conferees did not wear a mask. The test rendered a χ^2 (2) value of 6.933 which was significant with p=.031. Post hoc analysis with Wilcoxon signed-rank tests was then performed. It resulted that there was a significant difference between spatial audio without automatic head rotation and spatial audio with automatic head rotation (Z=-2.559, p=.010) with a moderate effect size (r=-0.395). These results indicate that in the condition where offline conferees did not wear surgical masks, spatial audio without head rotation led to an easier identification of the speaker in the conference compared to spatial audio with automatic head rotation. [with mask] A non-parametric Friedman's test of differences among repeated measures was conducted to compare the difficulty to identify the speaking conferee in the mono (M=3.14, SD=1.049), spatial (M=3.64, SD=1.122) and spatial with automatic head rotation (M=3.02, SD=1.024) auditory conditions, and in the visual condition where the conferees wore a mask. The test rendered a χ^2 (2) value of 8.918 which was significant with p=.012. Post hoc analysis with Wilcoxon signed-rank tests was then performed. It resulted that there was a significant difference between spatial audio without automatic head rotation land spatial audio with automatic head rotation (Z=-2.253, p=.024) with a moderate effect size (r=-0.348). These results indicate that in the condition where offline conferees wore surgical masks, spatial audio without head rotation led to an easier identification of the speaker in the conference than spatial audio with automatic head rotation.		No *** *** *** *** *** *** ***	No No
PostOverallComprehension	o Likert's scale (1 - 5)	Per participant	Higher	GLMM (ordinal)	3.726190	3.547619	3.821429	No	No	3.753968	0.864282	3.642857	0.864282	No	No		442- 332- 344-	No.	No
PostConcentration (Pairwise comparison doesn't have significant when we separate mask and no mask and hence not appear in the table)		Per participant	Higher	GLMM (ordinal)	2.726190	2.690476	3.095238	No	F(2,205) = 6.106 p = 0.003	2.428571	0.941731	3.246032	0.941731	No	F(2,205) = 60.939 p < 0.001	(Loic) [without mask] (this is without bonferroni correction) Post hoc analysis with Wilcoxon signed-rank tests was then performed. It resulted that there was a significant difference between spatial audio with automatic head rotation and mono audio (Z=-2.228, p=0.026) with a moderate effect (r=-0.344), and between spatial audio without automatic head rotation and spatial audio with automatic head rotation (Z=-1.962, p=.050) with a moderate effect size (r=-0.303).	treiant size of the state of th	No.	No
PostPresence	Likert's scale (1 - 5)	Per participant	Higher	GLMM (ordinal)	2.821429	2.928571	2.845238	No	No	3.095238	1.061535	2.634921	1.061535	No	F(2,205) = 19.400 p < 0.001		130-	No	No
PostAudioHelp	Likert's scale (1 - 5)	Per participant	Higher	GLMM (ordinal)	3.333333	3.452381	2.833333	No	F(2,205) = 11.214 p < 0.001	3.150794	1.051228	3.261905	1.051228	No	No	(Loic) [without mask] A non-parametric Friedman's test of differences among repeated measures was conducted to compare the helpfulness of the audio in the mono (M=3.40, SD=.939), spatial (M=3.29, SD=1.111) and spatial with automatic head rotation (M=2.76, SD=1.008) auditory conditions, and in the visual condition where the conferees did not wear a mask. The test rendered a χ^2 (2) value of 8.715 which was significant with p=.013. Post hoc analysis with Wilcoxon signed-rank tests was then performed. It resulted that there was a significant difference between mono and spatial audio with automatic head rotation (Z=-2.935, p=.003) with a moderate effect size (r=-0.453), and between spatial audio without automatic head rotation and spatial audio with automatic head rotation and spatial audio with automatic head rotation was effect size (r=-0.338). These results indicate that in the condition where offline conferees did not wear surgical masks, participants found that spatial audio with automatic head rotation was less helpful than mono or spatial audio without head rotation. [with mask] A non-parametric Friedman's test of differences among repeated measures was conducted to compare the helpfulness of the audio in the mono (M=3.26, SD=.885), spatial (M=3.62, SD=1.081) and spatial with automatic head rotation (M=2.90, SD=1.100) auditory conditions, and in the visual condition where the conferees wore a mask. The test rendered a χ^2 (2) value of 8.676 which was significant difference between spatial audio without automatic head rotation and spatial audio with automatic head rotation where offline conferees wore surgical masks, participants found that spatial audio with automatic head rotation wheat chead rotation was less helpful than spatial audio without head rotation was less helpful than spatial audio without head rotation.		No.	No
nb_fixation (per dialog duration, i.e. count/ms) (Pairwise comparison doesn't have significant when we separate mask and no mask and hence not appear in the table)		Per participant		GLMM (gamma)	0.002377	0.002364	0.002416	No	No	0.002395	0.000288	0.002377	0.000288	χ2(1, N = 252)= 8.0407,p = 0.004574	No	addio without flead fotation.	# mail # m # mail	χ2(2, N = 252)= 8.2709,p = 0.015995	F(2,205) = 5.020 p = 0.02354
mean_fix_duration (in ms) (Pairwise comparison doesn't have significant when we separate mask and no mask and hence not appear in the table)		Per participant	-	GLMM (gamma)	174.810926	175.499303	175.155535	No	No	176.866937	17.831948	173.443572	17.831948	No	F(1,205) = 9.8992 p = 0.0019		g 100	No	No
total_fix_duration	Computed for each video as the ratio of the sum of the fixations duration of each participant by the duration of the video in milliseconds in order to compare the means within the different conditions (in ms/ms)	Per participant		GLMM (normal) = LMM	0.417795	0.418002	0.424752	-	No	0.425530	0.077332	0.414836	0.077332		F(1,205) = 4.7183 p = 0.03099		1044.	maak to see the see that the se	No
mean_saccade_amplitud (Pairwise comparison doesn't have significant when we separate mask and no mask and hence not appear in the table)		Per participant		GLMM (gamma)	163.524391	163.098054	165.764959	No	No	156.262716	43.604398	171.995553	43.604398	χ2(1, N = 252)= 9.3530,p = 0.002226	F(1,205) = 20.1378 p < 0.001		Sound many Sound	No maak	No
mean_dissimilarity	Computed using the Dynamic Time Warping (DTW) algorithm with a custom cost function (assuming a cost of 1 when there is a difference between an element of the scanpath) on the scanpath of one participant to all the others for each video. We compute a mean similarity/dissimilarity with the resulting values for each participant and per video. The DTW algorithm works on timeseries data that might differ in length, which is the case with the obtained scanpaths as not every recording made by the participants led to the same number of fixation points. This algorithm aligns two timeseries and computes the similarity between them.		Lower	GLMM (normal) = LMM	116.511614	96.458188	98.893148		F(2,205) = 65.7189 p < 0.001	101.491289	16.388982	106.417344	16.388982		F(1,205) = 9.9882 p = 0.0018140	(Loic) There is a significant difference between mono (M=110,777, SD=16.920) and the two conditions with binaural sound: spatial (M=98.134, SD13.816), head (M=95.563, SD=14.348). The two conditions with binaural sounds lead to more similar scanpaths. As a scanpath is the sequence of fixation points, they represent the scanning strategy used by the participants over the whole video. If the scanpaths share similar sequences, they are then more similar. Thus, the obtained results indicate that in the conditions with spatial audio, the scanning strategies where more similar. One reason for this might be that the binaural produce a sequence of spatial cues that one might follow whereas in mono condition audio these spatial cues do not exist and thus the participant might follow different scanning strategies.	Stored Stored	mask ## to	F(2,205) = 7.8429 p < 0.001
ratio_aoi_transitions	Computed with the ROIs changes present in the scanpaths of each participant in each condition. A ROI change is identified in the scanpaths when two consecutives letter are not the same. The number of ROI changes in a given scanpath is equal to its length when abstracting the consecutive same letters. This number is then divided by the number of speech acts in the video of the corresponding condition to be able to compare the conditions between them. This ratio has to be done as the ROI changes are intrinsically dependant on the number of speech acts in		Lower	GLMM (normal) = LMM	2.381674	2.138522	2.075871	-	F(2,205) = 14.060 p < 0.001	1.965478	0.498773	2.4	0.498773		F(1,205) = 87.917 p < 0.001	(Loic) There is a significant difference between all the conditions. The spatial condition lead to more similarity between the scanpaths (M=94.783, SD=12.396), followed by head (M=102.223, SD=15.664), and finally by mono (M=122.246, SD=16.782). Similarly as before, the binaural audio induce a strategy that the participants can follow leading to less dissimilarity.	220	XX	F(2,205) = 14.644 p < 0.001
Mean dispersion per frame	Computed frame by frame as the mean of the Euclidean distances between all the fixations points present on the frame. All the videos have a frame rate of 25 frames per second, thus a duration of 40 milliseconds. The fixation points used for computing the dispersion were those present in the interval of time of the frame. The dispersion allows to compare the variability between the fixations points: a low dispersion value means that the fixations points are more clustered. There is a limitation to this metric as stated by Coutrot et al. [5]: the gaze and thus the fixation points, could be clustered but on different region of interests, resulting in an increase in dispersion even if the points are clustered.		Lower	Linear model (not so sure	2) 14.564533	14.525594	14.515028	No		14.481805	6.369241	14.590462	6.451730	No			Seord Sound	No No	
Mean dispersion per window at transitions		Per dialog	Lower	Linear model (not so sure	e) 15.705252	15.643907	15.678048	No		15.665718	2.654992	15.686313	2.728477	No			Ti-	No.	
Mean dispersion per window at interruptions	Same as mean dispersion but the mean is averaged on the window of frames contain a interruption	Per dialog	Lower	Linear model (not so sure	15.658549	16.126366	4.238106	No		15.898706	4.108529	16.054101	4.249273	No		(Loic) [without mask] There is a significant difference of mean dispersion between all auditory conditions, with small effect sizes. Spatial condition has the lowest dispersion between fixation points (M= 40.669, SD=27.777), followed by mono (M = 44.097, SD=29.139) and finally by head (M= 45.707, SD=29.733). This indicates that the spatial conditions generate more clustered fixation points whereas the mono and head generate less clustered fixation points.	Thomas Sand	No.	