/

Appendix 1: Statistical analysis results for Study I

Table 1: *Statistical results for the tests performed on understandability assessment measures compared within participants of ATL-Viz group.*

Dependent variable	Statistics	
	ATL-Viz	ATL-Viz Control
Task completion time (seconds)	$Z = -3.40$, $p < .000$, $p_{\text{adjusted}} < .006$	
	$Mdn = 288.9$, $IQR = 198.4$	$Mdn = 456.7$, $IQR = 253.0$
Number of errors	$Z = -3.94$, $p < .000$, $p_{\text{adjusted}} < .000$	
	$Mdn = 15.0$, $IQR = 10.0$	$Mdn = 50.0$, $IQR = 22.5$

Table 2: *Statistical results for the tests performed on understandability assessment measures compared within participants of RAD-Viz group*

Dependent Variable	Statistics	
	RAD-Viz	RAD-Viz Control
Task completion time (seconds)	$Z = -3.64$, $p < .000$, $p_{\text{adjusted}} < .000$	
	$Mdn = 265.0$, $IQR = 97.4$	$Mdn = 498.2$, $IQR = 212.2$
Number of errors (percent)	$Z = -3.26$, $p = .001$, $p_{\text{adjusted}} < .000$	
	$Mdn = 10.0$, $IQR = 30.0$	$Mdn = 40.0$, $IQR = 27.5$

Table 3: *Statistical results for the tests performed on understandability assessment measures compared between participants of ATL-Viz and RAD-Viz groups*

Dependent Variable	Statistics			
	ATL-Viz	RAD-Viz	ATL-Viz (control)	RAD-Viz (control)
Task completion time (seconds)	$U = 1.286$, $p = .49$		$U = -.292$, $p = .381$	
Number of errors(percent)	$U = .892$, $p = .763$		$U = 1.46$, $p = .381$	

Appendix 2: Statistical analysis results for Study II

Table 4 presents the results for within-participant of each VA interface group (ATL-Viz and RAD-Viz) performed in Study II (Section 4.5 of the paper). For each group comparison was made between four conditions. For example, for ATL-Viz the conditions were: low complexity as visualised on ATL-Viz, low complexity as visualised on the control display, high complexity as visualised on ATL-Viz and high complexity as visualised on the control display. Table 5 presents the statistical results for comparison between the two VA interface groups (ATL-Viz and RAD-Viz). For each dependent variable, two display conditions were compared once for the low-complexity scenario and once for the high-complexity scenario. Choice of statistical tests were made based on the number of conditions and data type. Median and IQR values are reported whenever a significant effect was observed. Coloured dots mark the conditions whose effect was significant.

Table 4: Results for statistical tests performed on decision-making measures compared within participants of each VA group tested

Dependent Variable	statistics			
	ATL-Viz	ATL-Viz Control	RAD-Viz	RAD-Viz Control
Number of clicks made regarding conflicts on the radar screen	$\chi^2 = 43.63, p < .000$		$\chi^2 = 47.48, p < .000$	
	BS: 0 (0)●● CM: 0 (0)●●	BS: 17 (11.0)●● CM: 18.5 (10.75)●●	BS: 0 (0)●● CM: 0 (0)●●	BS: 17 (11.0)●● CM: 15.5 (10.0)●●
Number of conflicts solved on the radar screen	$\chi^2 = 54.0, p < .000$		$\chi^2 = 50.65, p < .000$	
	BS: 0 (0)●● 0 (0)●●	BS: 5 (0)●● CM: 5 (0)●●	BS: 0 (0)●● 0 (0)●●	BS: 5 (0)●● CM: 5 (0)●●
Number of conflicts solved with the order of urgency	$\chi^2 = 2.45, p = .48$		$\chi^2 = 5.77, p = .12$	
Number of ROCD & HDG resolutions made on the radar screen	$\chi^2 = 51.8, p < .000$		$\chi^2 = 49.9, p < .000$	
	BS: 0.0 (0.0)●● CM: 0.0 (0.0)●●	BS: 3 (1.2)●● CM: 4 (2.2)●●	BS: 0 (0)●● CM: 0 (0)●●	BS: 4 (1.0)●● CM: 4 (1.7)●●
Time to first interaction (seconds)	$\chi^2 = 50.7, p < .000$		$\chi^2 = 28.0, p < .000$	
	BS: 3.0 (3.0)●● CM: 2.0 (1.2)●●	BS: 13.5 (8.7)●● CM: 17.5 (8.7)●●	BS: 4.5 (3.7)●● CM: 5.0 (2.0)●●	BS: 11.5 (8.0)●● CM: 15 (8.0)●●
Number of conflict ignored	$\chi^2 = 12.7, p = .005$		$\chi^2 = 9.43, p = .024$	
	BS: 0.0 (0.0)● CM: 0.0 (0.0)	BS: 0.0 (1.0)● CM: 0.0 (1.0)	BS: 0.0 (0.0) CM: 0.0 (0.0)	BS: 0.0 (0.0) CM: 0.0 (0.0)
				Continued on next page

Table 4 – continued from previous page

Dependent Variable	statistics			
	ATL-Viz	ATL-Viz Control	RAD-Viz	RAD-Viz Control
Time to have CD&R tasks accomplished (seconds)	$\chi^2 = 30.7, p < .000$		$\chi^2 = 19.12, p < .000$	
	BS: 83.5 (42.0)●●	BS: 143.5 (55.2)●●	BS: 93.0 (38.7)●	BS: 111.5 (60.7)
	CM: 74.5 (42.0)●●	CM: 165.0 (55.2)●●	CM: 100.0 (81.0)●	CM: 157.0 (60.7)●●
decision-making duration (seconds)	Conflict A: $\chi^2 = 8.66, p = .034$ $p_{\text{adjusted}} > .05$		$\chi^2 = 1.22, p = .75$	
	Conflict B: $\chi^2 = 16.2, p = .001$		$\chi^2 = 4.0, p = .26$	
	BS: 14.0 (8.7)●●	BS: 19.5 (16.2)●●		
	CM: 11.5 (6.2)●●	CM: 26.5 (16.2)●●		
	Conflict C: $\chi^2 = 22.6, p < .000$		$\chi^2 = 23.0, p < .000$	
	BS: 13.5 (6.0)●●	BS: 27.5 (16.5)●●	BS: 16.5 (11.0)●	BS: 18.5 (18.5)
	CM: 11.0 (12.2)●●	CM: 30.0 (16.5)●●	CM: 16.5 (10.7)●	CM: 35.5 (18.5)●●
	Conflict D: $\chi^2 = 10.4, p = .02$		$\chi^2 = 8.14, p = .04$	
	BS: 13.5 (17.0)	BS: 22.0 (15.5)	BS: 16.0 (14.7)●	BS: 13.5 (18.5)●
	CM: 14.0 (12.0)●	CM: 25.0 (15.5)●	CM: 13.5 (9.0)	CM: 32.0 (18.5)●●
	Conflict E: $\chi^2 = 16.2, p = .001$ $p_{\text{adjusted}} > .05$		$\chi^2 = 11.66, p = .008$ $p_{\text{adjusted}} > .05$	
Resolution strategy	Conflict A $p_{(BS)} = .39, p_{(CM)} = .09$		$p_{(BS)} = .93, p_{(CM)} = .52$	
	Conflict B $p_{(BS)} = 1.0, p_{(CM)} = 1.0$		$p_{(BS)} = .93, p_{(CM)} = .78$	
	Conflict C $p_{(BS)} = .97, p_{(CM)} = .98$		$p_{(BS)} = .51, p_{(CM)} = .51$	
	Conflict D $p_{(BS)} = 1.0, p_{(CM)} = 1.0$		$p_{(BS)} = .70, p_{(CM)} = .40$	
	Conflict E $p_{(BS)} = .41, p_{(CM)} = .59$		$p_{(BS)} = .14, p_{(CM)} = .16$	
Workload	$\chi^2 = 19.84, p < .000$ $p_{\text{adjusted}} > .05$		$\chi^2 = 12.77, p = .005$ $p_{\text{adjusted}} > .05$	

BS stands for the baseline scenario (low complexity) and CM stands for the complex scenario (high complexity).

ROCD stands for rate of climb or descent.

HDG stands for heading.

CD&R stands for conflict detection and resolution.

Table 5: Results for statistical tests performed on decision-making measures compared between participants of ATL-Viz and RAD-Viz groups

Dependent Variable	statistics			
	ATL-Viz	RAD-Viz	ATL-Viz Control	RAD-Viz Control
Number of clicks made regarding conflicts on the radar screen	$U_{(BS)} = .51, p_{(BS)} = .93$ $U_{(CM)} = .77, p_{(CM)} = .79$		$U_{(BS)} = .76, p_{(BS)} = .87$ $U_{(CM)} = 1.31, p_{(CM)} = .47$	
Number of conflicts solved on the radar screen	$U_{(BS)} = .80, p_{(BS)} = .73$ $U_{(CM)} = 1.07, p_{(CM)} = .38$		Not applicable	
Number of conflicts solved with the order of urgency	$U_{(BS)} = 1.84, p_{(BS)} = .11$ $U_{(CM)} = 1.90, p_{(BS)} = .07$		$U_{(BS)} = 1.18, p_{(BS)} = .46$ $U_{(CM)} = -.19, p_{(CM)} = .19$	
Number of ROCD & HDG resolutions made on the radar screen	$U_{(BS)} = .58, p_{(BS)} = 1.0$ $U_{(CM)} = .58, p_{(CM)} = 1.0$		$U_{(BS)} = -.69, p_{(BS)} = .19$ $U_{(CM)} = .57, p_{(CM)} = 1.0$	
Time to first interaction (seconds)	$U_{(BS)} = -1.4, p_{(BS)} = .049$ $U_{(CM)} = -3.3, p_{(CM)} < .000$ BS: 3.0 (3.0)● CM: 2.0 (1.2)●	BS: 4.5 (3.7)● CM: 5.0 (2.0)●	$U_{(BS)} = 1.20, p_{(BS)} = .55$ $U_{(CM)} = .86, p_{(CM)} = .79$	
Mouse hover duration over glyph (seconds)	Conflict A: $U_{(BS)} = -.79, p_{(BS)} = .17$ $U_{(CM)} = .15, p_{(CM)} = .67$		Not applicable	
	Conflict B: $U_{(BS)} = -.57, p_{(BS)} = .25$ $U_{(CM)} = -1.17, p_{(CM)} = .08$ BS: 8.5 (5.7)	BS: 12.0 (9.7)	Not applicable	
	Conflict C: $U_{(BS)} = -1.72, p_{(BS)} = .02$ $U_{(CM)} = -.96, p_{(CM)} = .12$		Not applicable	
	Conflict D: $U_{(BS)} = -.70, p_{(BS)} = .20$ $U_{(CM)} = -.91, p_{(CM)} = .14$		Not applicable	
	Conflict E: $U_{(BS)} = -.31, p_{(BS)} = .38$ $U_{(CM)} = -.70, p_{(CM)} = .20$		Not applicable	
	Continued on next page			

Table 5 – continued from previous page

Dependent Variable	statistics			
	ATL-Viz	RAD-Viz	ATL-Viz Control	RAD-Viz Control
Total mouse hover duration over glyphs (seconds)	$U_{(BS)} = -1.20, p_{(BS)} = .08$ $U_{(CM)} = -1.14, p_{(CM)} = .09$		Not applicable	
decision-making duration (seconds)	Conflict A: $U_{(BS)} = -.38, p_{(BS)} = .34$ $U_{(CM)} = .17, p_{(CM)} = .69$		$U_{(BS)} = .54, p_{(BS)} = .98$ $U_{(CM)} = 1.7, p_{(CM)} = .27$	
	Conflict B: $U_{(BS)} = -.83, p_{(BS)} = .16$ $U_{(CM)} = -1.0, p_{(CM)} = .11$		$U_{(BS)} = 1.0, p_{(BS)} = .65$ $U_{(CM)} = 1.37, p_{(CM)} = .44$	
	Conflict C: $U_{(BS)} = -.26, p_{(BS)} = .40$ $U_{(CM)} = -1.1, p_{(CM)} = .09$		$U_{(BS)} = 1.4, p_{(BS)} = .45$ $U_{(CM)} = -.80, p_{(CM)} = .17$	
	Conflict D: $U_{(BS)} = .35, p_{(BS)} = .83$ $U_{(CM)} = .32, p_{(CM)} = .80$		$U_{(BS)} = 1.9, p_{(BS)} = .19$ $U_{(CM)} = -.10, p_{(CM)} = .50$	
	Conflict E: $U_{(BS)} = .69, p_{(BS)} = .93$ $U_{(CM)} = -.44, p_{(CM)} = .31$		$U_{(BS)} = 2.05, p_{(BS)} = .15$ $U_{(CM)} = -.31, p_{(CM)} = .38$	
Time to have CD&R tasks accomplished (seconds)	$U_{(BS)} = -.53, p_{(BS)} = .27$ $U_{(CM)} = -.77, p_{(CM)} = .18$		$U_{(BS)} = 1.97, p_{(BS)} = .17$ $U_{(CM)} = .34, p_{(CM)} = .81$	
Resolution strategies	Conflict A $p_{(BS)} = .31, p_{(CM)} = .13$		$p_{(BS)} = .76, p_{(CM)} = .39$	
	Conflict B $p_{(BS)} = .70, p_{(CM)} = .30$		$p_{(BS)} = 1.0, p_{(CM)} = .92$	
	Conflict C $p_{(BS)} = .30, p_{(CM)} = .81$		$p_{(BS)} = .95, p_{(CM)} = 1.0$	
	Conflict D $p_{(BS)} = .99, p_{(CM)} = .99$		$p_{(BS)} = .32, p_{(CM)} = .77$	
	Conflict E $p_{(BS)} = 1.0, p_{(CM)} = .99$		$p_{(BS)} = .95, p_{(CM)} = .63$	
Workload	$U_{(BS)} = -.10, p_{(BS)} = .50$ $U_{(CM)} = -.36, p_{(CM)} = .35$		$U_{(BS)} = -.31, p_{(BS)} = .38$ $U_{(CM)} = -.23, p_{(CM)} = .42$	

Appendix 3: Statistical analysis results for comparing dependent measures between ATCos and Novices

Table 6: Results of two-sided Mann-Whitney U test performed on decision-making (study II) measures compared between ■novices and ■ATCos. Median and IQR values are reported whenever the effect was significant.

Dependent Variable	Traffic complexity	statistics			
		ATL-Viz	RAD-Viz	ATL-Viz Control	RAD-Viz Control
Number of clicks made regarding conflicts on the radar screen	low	$U = 3.95, p = .161$	$U = 3.60, p = .074$	$U = 6.44, p = .174$	$U = 4.62, p = .676$
	high	$U = 4.06, p = .209$	$U = 3.84, p = .139$	$U = 6.67, p = 113$	$U = 4.92, p = .911$
Number of conflicts solved on the radar screen	low	$U = 4.12, p = .209$	$U = 4.92, p = .839$	Not applicable	Not applicable
	high	$U = 4.09, p = .196$	$U = 4.73, p = .503$	Not applicable	Not applicable
Number of conflicts solved with the order of urgency	low	$U = 5.0, p = .96$	$U = 5.0, p = 1.0$	$U = 4.78, p = .731$	$U = 4.15, p = .271$
	high	$U = 5.23, p = .799$	$U = 5.92, p = .334$	$U = 4.81, p = .743$	$U = 5.23, p = .799$
Time to first interaction (seconds)	low	$U = 5.11, p = .97$	$U = 3.42, p = .36$	$U = 3.37, p = .10$	$U = 1.72, p = .009$ ■11.5(8.0) ■22.0(19.0)
	high	$U = 3.04, p = .043$ ■2.0(1.25) ■5.0(3.5)	$U = 3.75, p = .56$	$U = 3.35, p = .09$	$U = 3.30, p = .30$
Total Mouse hover duration over glyphs (seconds)	low	$U = 4.09, p = .34$	$U = 2.60, p = .08$	Not applicable	Not applicable
	high	$U = 3.21, p = .06$ ■41.5(28.25) ■54.0(81.5)	$U = 1.69, p = .008$ ■48.5(40.5) ■112.0(71.5)	Not applicable	Not applicable
Number of ROCD&HDG resolutions made on the radar screen	low	$U = 3.40, p = .003$ ■0.0(0.0) ■0.0(2.0)	$U = 3.81, p = .13$	$U = 6.36, p = .188$	$U = 6.08, p = .07$
	high	$U = 3.40, p = .003$ ■0.0(0.0) ■0.0(1.0)	$U = 1.63, p < .000$ ■0.0(0.0) ■2.0(3.0)	$U = 5.95, p = .376$	$U = 8.14, p < .000$ ■4.0(1.75) ■0.0(0.0)
Number of conflicts ignored	low	$U = 6.06, p = 1.0$	$U = 4.36, p = 1.0$	$U = 6.42, p = .08$	$U = 4.78, p = .40$
	high	$U = 5.26, p = .61$	$U = 4.36, p = 1.0$	$U = 6.03, p = .16$	$U = 5.20, p = .20$
Time to have CD&R tasks accomplished (seconds)	low	$U = 2.02, p = .002$ ■83.5(42.0) ■196(141.5)	$U = 1.09, p = .001$ ■93.0(38.75) ■218.0(135.5)	$U = 4.18, p = .39$	$U = 1.84, p = .013$ ■111.5(60.75) ■176.0(62.5)

Table 6 – continued from previous page

Dependent Variable	Traffic complexity	statistics			
		ATL-Viz	RAD-Viz	ATL-Viz Control	RAD-Viz Control
	high	$U = 1.71, p < .001$ ■74.5(42.0) ■188.0(136.5)	$U = 1.66, p = .007$ ■100.0(81.0) ■185.0(45.0)	$U = 4.65, p = .70$	$U = 2.12, p = .03$ ■157.0(74.5) ■201.0(66.0)
decision-making duration (seconds)	low	A: $U = 3.48, p = .12$ B: $U = 2.32, p = .006$ ■14.0(8.75) ■30.0(38.0) C: $U = 3.07, p = .049$ ■13.5(6.0) ■34.0(20.5) D: $U = 2.30, p = .006$ ■13.5(17.0) ■42.0(36.5) E: $U = 2.32, p = .006$ ■15.0(11.25) ■39.0(34.5)	A: $U = 3.24, p = .276$ B: $U = 1.57, p = .006$ ■14.5(11.5) ■48.0(31.0) C: $U = 1.99, p = .019$ ■16.5(11.0) ■26.0(18.0) D: $U = .877, p < .001$ ■16.0(14.75) ■41.0(20.0) E: $U = 2.18, p = .031$ ■15.0(20.0) ■32.0(14.0)	A: $U = 4.45, p = .56$ B: $U = 3.71, p = .18$ C: $U = 4.43, p = .54$ D: $U = 2.90, p < .033$ ■22.0(15.5) ■33.0(10.0) E: $U = 4.70, p = .74$	A: $U = 2.54, p = .073$ B: $U = 1.90, p = .015$ ■19.5(15.2) ■42.0(14.5) C: $U = 2.36, p = .049$ ■18.5(18.5) ■33.0(17.0) D: $U = 1.87, p = .014$ ■13.5(18.5) ■33.0(20.5) E: $U = 3.45, p = .38$
	high	A: $U = 3.65, p = .16$ B: $U = 1.88, p = .001$ ■11.5(6.25) ■38.0(18.0) C: $U = 2.32, p = .006$ ■11.0(12.25) ■29.0(29.0) D: $U = 1.66, p < .001$ ■14.0(12.0) ■48.0(41.0) E: $U = 2.90, p = .033$ ■14.0(13.75) ■30.0(28.5)	A: $U = 2.54, p = .07$ B: $U = 2.05, p = .02$ ■15.0(15.0) ■29.0(5.0) C: $U = 1.24, p = .002$ ■16.5(10.75) ■29.0(11.5) D: $U = 1.94, p = .016$ ■13.5(9.0) ■46.0(39.0) E: $U = 3.51, p = .413$	A: $U = 5.09, p = 1.0$ B: $U = 4.5, p = .59$ C: $U = 3.62, p = .16$ D: $U = 3.48, p = .12$ E: $U = 5.56, p = .64$	A: $U = 2.21, p = .03$ ■26.5(10.25) ■37.0(16.0) B: $U = 1.81, p = .01$ ■22.0(12.5) ■36.0(28.5) C: $U = 3.12, p = .22$ D: $U = 3.48, p = .39$ E: $U = 4.33, p = 1.0$
Resolution strategies	low	A: $p = .43$ B: $p = .86$ C: $p = 1.0$ D: $p = 1.0$ E: $p = .05$	A: $p = .06$ B: $p = .35$ C: $p = .37$ D: $p = .51$ E: $p = .65$	A: $p = .23$ B: $p = .28$ C: $p = .61$ D: $p = .86$ E: $p = .29$	A: $p = .001$ $p_{adjusted} = .28$ B: $p = .14$ C: $p = .75$ D: $p = .56$ E: $p = .02$ $p_{adjusted} = .19$
	high	A: $p = .71$ B: $p = .60$ C: $p = .18$ D: $p = .84$ E: $p = .05$	A: $p = .18$ B: $p = .10$ C: $p = .33$ D: $p = 1.0$ E: $p = .82$	A: $p = .70$ B: $p = .15$ C: $p = .03$ $p_{adjusted} = .22$ D: $p = .85$ E: $p = .56$	A: $p = .003$ $p_{adjusted} = .46$ B: $p = .34$ C: $p = .03$ $p_{adjusted} = .37$ D: $p = .55$ E: $p = .14$
Workload	low	$U = 4.62, p = .677$	$U = 4.81, p = .578$	$U = 6.197, p = .264$	$U = 7.331, p = .023$ ■50.0(27.5) ■40(22.5)
	high	$U = 4.04, p = .316$	$U = 3.84, p = .232$	$U = 6.64, p = .117$	$U = 7.30, p = .026$ ■69.0(21.25) ■50(32.5)

Table 7: *Statistical results for the tests performed on understandability assessment measures compared between ■novices and ■ATCos. Median and IQR values are reported whenever the effect was significant.*

Dependent Variable	Statistics			
	ATL-Viz	RAD-Viz	ATL-Viz (control)	RAD-Viz (control)
Task completion time (seconds)	$U = 5.28, p = .85$	$U = 1.51, p = .003$ ■265.3(97.4) ■453.3(105.2)	$U = 3.96, p = .29$	$U = 3.63, p = .49$
Number of errors	$U = 7.88, p = .003$ ■15.0(10.0) ■0.0(0.0)	$U = 5.08, p = .46$	$U = 6.56, p = .13$	$U = 5.39, p = .31$

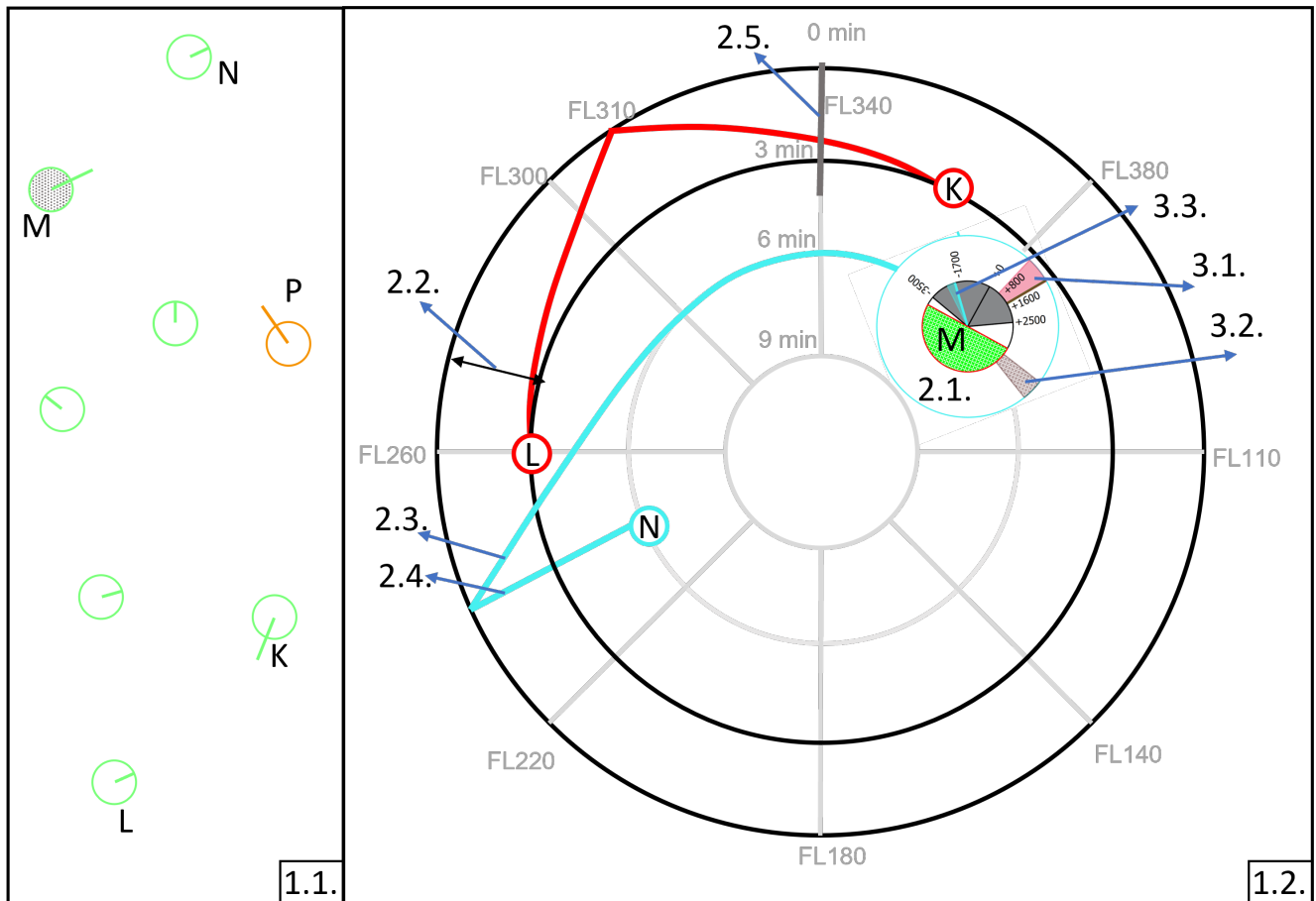**Appendix 4: Description of RAD-Viz interface**

Figure 1: *Schematic representation of RAD-Viz interface.*

Figure 1 depicts a schematic description of RAD-Viz interface showing the same traffic scenario depicted on ATL-Viz schematic description in Section 3.2. of the paper. Similar to ATL-Viz, RAD-Viz interface consists of a radar screen (1.1.) and the time-altitude display (1.2.). The radar screen is identical on both interfaces. However on RAD-Viz time-altitude display, aircraft glyph are mapped on inverted axes of the polar graph compared to ATL-Viz. On RAD-Viz, the flight level information is depicted on the angular axis and time remaining to conflict is depicted on the radial axis. The numbered visual items correspond to the structural properties obtained from the functional layers of the WDA (see sections 3.1. and 3.2. in the paper). 2.1. depicts the glyph and its visual components (heading and ROCD solution spaces). The outermost black circle indicates separation loss occurrence (zero time). The other black circle, points the most imminent conflict and expands towards the zero time reference circle as time passes. 2.2. indicates time to conflict. 2.3. indicates vertical trajectory profile and 2.4. indicates altitude criteria to avoid for resolving the current conflict. 2.5. indicates altitude criteria to avoid potential conflicts. Upon hovering mouse over the glyph, the solution spaces are shown in details (as depicted in the figure). 3.1. indicates heading criteria to avoid current conflicts, 3.2 indicates heading criteria to avoid potential conflicts. 3.3 indicates ROCD criteria to avoid current conflicts. Various heading and rate of climb values can be explored by right clicking on the outer and inner circles of the glyph respectively. To apply the changes, a confirmation box will appear upon left clicking on the outer circle where the selected values can be confirmed. As can be seen from the figure, the time-altitude display shows four out of eight aircraft shown on the radar screen are in conflict. Aircraft K and L will lose separation in 3 minutes at flight level(FL) 310. Aircraft M and N will lose separation at FL240. Aircraft M is selected (indicated by the green half circle inside the glyph. If M is sent to FL 340 or its heading is changed to the patterned section (3.2. on the glyph), it will have conflict with aircraft P.