Model-Based Analysis of Two-Alternative Decision Errors in a Videopanorama-Based Remote Tower Work Position

Norbert Fürstenau, Monika Mittendorf, and Maik Friedrich

German Aerospace Center, Inst. of Flight Guidance, Braunschweig, Germany
Norbert.fuerstenau@dlr.de

Abstract. Initial analysis of a first Remote Control Tower (RTO) field test with an experimental videopanorama system [1] [2] under quasi operational conditions has shown performance deficits quantified by two-alternative aircraft maneuver discrimination tasks [3]. RTO-controller working position (CWP-) performance was compared with that one of the conventional tower-CWP with direct out-of-windows view by means of simultaneous aircraft maneuver observations at both operator positions, and it was quantified using discriminability d' and Bayes inference. Here we present an extended data analysis using nonparametric discriminability A and we discuss the RTO performance deficit in terms of the information processing (IP) theory of Hendy et al. [4]. As initial working hypothesis this leads to the concept of time pressure (TP) as one major source of the measured response errors. We expect the RTO-performance deficits to decrease with the introduction of certain automation features to reduce time pressure and improve the usability of the videopanorama system. A fit of the experimental data with a modified error vs. TP function provides some evidence in support of the IP/TP-hypothesis, however more specifically designed experiments are required for obtaining sufficient confidence.

Keywords: Remote Tower, videopanorama, field testing, flight maneuvers, two-alternative decisions, signal detection theory, information processing theory, time pressure.

1 Introduction

Since about ten years remote control of low traffic airports (Remote Tower Operation, RTO) has emerged as a new paradigm to reduce cost of air traffic control [1]. It was suggested that technology may remove the need for local control towers [5]. Controllers could visually supervise airports from remote locations by videolinks, allowing them to monitor many airports from a remote tower center (RTC) [2]. It is clear from controller interviews that usually numerous out-the-window visual features are used for control purposes [6]. In fact, these visual features go beyond those required by regulators and ANSP's (air navigation service providers) which typically include only aircraft detection, recognition, and identification [7]. Potentially important additional visual features identified by controllers in interviews involve subtle aircraft motion. In

D. Harris (Ed.): EPCE 2014, LNAI 8532, pp. 143–154, 2014.

fact, the dynamic visual requirements for many aerospace tasks have been studied, but most attention has been paid to pilot vision (e.g. [8]). In this work we investigate a group of visual cues derived from flight maneuvers within the range of observability in the control zone. They might be indicative of aircraft status and pilots situational awareness which is important with the higher volume of VFR traffic in the vicinity of small airports.

These considerations led to the design of the present validation experiment within the DLR project RAiCe (Remote Airport traffic Control Center, 2008 – 2012). The field test was realized within a DLR - DFS (German ANSP) Remote Airport Cooperation. Specifically dual-choice decision tasks (the subset of “Safety related maneuvers” in [9]) were used for quantifying the performance difference between the standard control tower work environment (TWR-CWP) and the new RTO controller working position (RTO-CWP) based on objective measures from signal detection theory (SDT)[10] and Bayes inference [3]. Here we confirm these preliminary results by additional data evaluation using the nonparametric discriminability index A [11] and present a new model-based analysis in terms of the information processing/time pressure (IP/TP-) theory of Hendy et al.[4] for comparing the measured performance deficit of the RTO-CWP with the predictions of a theoretical error model.

Experimental methods are reviewed in section 2 followed by the results in section 3 (response times, Hit and False Alarm rates). Using these data in section 4 nonparametric discriminability coefficients are calculated and error rates are fitted with a IP/TP based model. We finish with a conclusion and outlook in section 5.

2 Methods

In what follows we review the experimental design with two-alternative decision tasks as part of the remote tower validation experiment and present additional details relevant for the IP-theory based analysis. Further details of the full passive shadow mode validation trial are reported in [9].

2.1 Participants

Eight tower controllers (ATCO’s) from DFS were recruited as volunteer participants for the experiment. The average age was 30 (stdev 12) years with 10 (stdev. 10) years of work experience, and they came from different small and medium airports. They took part at the experiment during normal working hours and received no extra payment. They were divided into 4 experimental pairs for simultaneously staffing the control tower (TWR-CWP) and the RTO-CWP.

2.2 Experimental Environment and Conditions

The experiment was performed as passive shadow mode test under quasi operational conditions on the four days July 17 – 20 2012. The remote tower system used in the present experiment was located at the DFS-operated Erfurt-Weimar (EDDE) control tower. It was an improved version of the RTO-experimental testbed at Braunschweig airport which was in use since 2004 for initial verification and validation trials [1][2]. Figures 1 show the sensor system and the RTO-CWP with 200° - videopanorama and operator console based on a reconstructed far view with five HD-format 40”-displays (892 x 504 mm, 1920x1080 pixel, pixel distance = 0.47 mm). A separate monitor (left table side) displays the pan-tilt zoom camera which is controlled via a pen-input interaction display with virtual joystick. Twelve preset viewing directions and four zoom factors Z = 2, 4, 8, 16 (viewing angles 26° - 3°) could be selected. Additional monitors include (middle row from right to left): flight plan data, approach radar and weather display. In contrast to the experimental TWR-CWP the RTO-CWP was on the TWR ground floor in a separate room without visual contact to the airfield. The TWR-CWP was located close to the operational ATCO, but they were instructed not to communicate with each other.

Fig. 1. Remote tower installation with 200°-panorama and pan-tilt zoom camera sensor system at the tower roof (left photo), and operators workplace with 40”-HD-format displays (right)

Pre-defined flight maneuvers were generated with a DLR DO228 twin turboprop engine test aircraft (D-CODE, length 15.03 m, body height x width 1.8 x 1.6 m, wing span 16.97 m, wheel diam. 0.65 m) performing aerodrome circling.

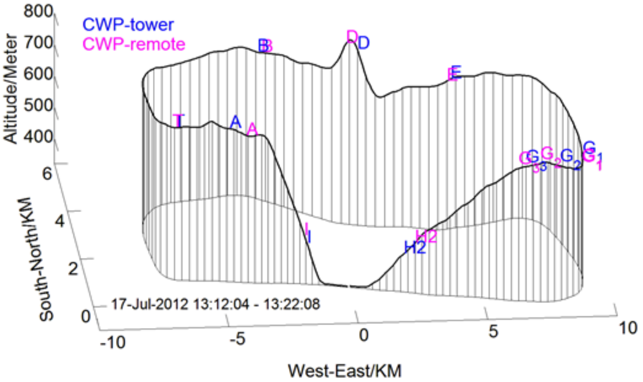

Fig. 2. DCODE trajectory measured with on-board satellite navigation. Letters indicate event positions with logged answers of TWR-CWP (blue) and RTO-CWP (red) operators to simultaneous task oriented on-line questioning. Distance between vertical lines = 5 s, projected to trajectory altitude minimum at ca. 350 m (sea level). Tower position (0, 0), height = 30 m.

The set of five well defined flight-maneuvers as stimuli for decision tasks at given positions within the EDDE control zone is indicated in Fig.2 with a 3D plot of the logged on board GPS trajectory. Trajectory minimum altitude represents a runway overflight at about 30 m above ground. The two types of maneuver-stimuli could be observed either visually-only (e.g. landing gear down) or visually and by radar (altitude change). During the experiment sometimes additional low volume normal traffic took place which now and then lead to delays in the traffic circle. Average duration of a full circle (= one run) was ca. 10 min yielding typically 140 min of experiment duration per participant pair for the nominally 14 full circles.

Radio communication between D-CODE pilots and flight engineer and the experimenter at the tower was realized with a separate research frequency in addition to the standard A/C–TWR radio channel. The available time for participant's responses to decision tasks was limited so that correct, incorrect and non-answers were possible.

2.3 Experimental Design and Task

Based on the fixed set of evaluated two-alternative events during a single circle (A, D, G_1, G_2, $H_{1,2,3}$) the concrete event situation (stimulus alternative S1 = maneuver, S2 = no maneuver) for decision making were switched statistically between runs according to two mirrored scenarios with different task sequences. In this way during the 14 circles per experiment / participant for each event about 7 answers per event alternative and per participant were obtained for the analysis (TWR and RTO CWP condition as independent variables). The within subjects design (TWR vs. RTO-CWP) yields about N = 56 answers and non-answers for averaging with the 8 participants.

During one aerodrome circling the two participants at the TWR and RTO-CWP had to simultaneously answer 19 different types of questions concerning the D-CODE maneuvers (events), object detection, and weather status. For analysis with discriminability index A and IP/TP theory based error model the following subset of 5 of the 9 event related questions is evaluated with regard to hit- and false alarm rates (in brackets: maximum response time Ta = interval until next task / question): 1. Does A/C perform repeated bank angle changes? (event position A; Ta = 20 s), 2. altitude variation ? (by 300 ft, event pos. D; 20 s), 3. landing light-off? Report status, event pos. G_1; 180 sec: switching-off not observable), 4. A/C on glide path?, event pos. G_2; 90 s; 5. Landing gear-down?; Report during final approach;event pos. $H_{1,2,3}$, distance 1.5, 1, 0.5 km; 10 s). A subjective certainty rating on a 5-point scale was not evaluated for the present analysis.

Every pair of participants had to complete two experimental trials. For the first trial (duration approximately 140 min) in the morning till noon, the participants were randomly assigned to one of the two CWP's. Positions were changed for the second trial in the afternoon. The number of airport circles corresponding to the number of decisions for each specific event was between 10 and 14. With two alternative stimuli (S_1, S_2) per task each pair of participants was exposed to 5 – 7 stimuli of the same kind in random succession for each event under TWR and the same number under RTO conditions. The present data analysis is focused on deriving objective measures for the two-alternative decision tasks. Additional data evaluation was presented in a parallel paper [9] addressing performance (answers given, response times, and sources of information) and subjective measures (debriefing, questionnaires).

3 Results

The response matrices of conditional probabilities $p(y|S_1)$ = hit rate H, $p(n|S_1)$ = misses M, $p(n|S_2)$ = correct rejections CR, $p(y|S_2)$ = false alarms FA, for the two alternative situations (stimuli), S1, S2, structure the results of each of the five events. Because participant's responses to event related questions were allowed to be positive, negative, and non-answers (no decision during the available time Ta), we analyse two types of response matrices: a) (optimistic) neglecting non-answers, b) (pessimistic) interpreting non-answers as false decisions (M or FA). In this way we obtain for each of the decision tasks an optimistic and a pessimistic estimate with regard to decision errors. The percentage correct analysis in [9] and the preliminary SDT and Bayes inference analysis [3] had shown that neglection of non-answers suggested no significant performance difference between TWR-CWP and RTO-CWP. The interpretation of the non-answers as erroneous responses appears to be justified due to increased uncertainty about the correct answer resulting in hesitation to respond at all because tower controllers work ethics requires decision making with high certainty. Figure 3 shows the statistics of non-answers, separated for the TWR-CWP and RTO-CWP condition.

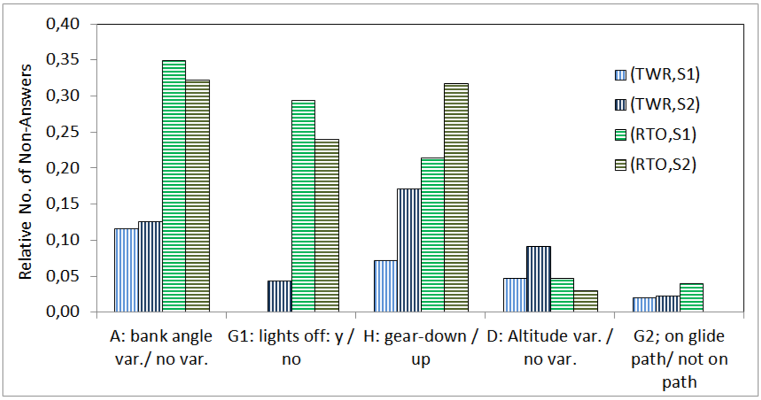

Fig. 3. Relative number of non-answers for the five analyzed decision tasks, separated for the two conditions TWR-CWP(left two columns, blue, vertical lines), RTO-CWP(right columns, green, horizontal lines), normalized with regard to the two respective alternative situations S1 (flight maneuver / stimulus, light colour), S2 (no flight maneuver / stimulus, dark colour))

Within the theoretical framework of SDT the two alternative stimuli S_1, S_2 for each event define independent statistical variables. Each set of decisions of a single subject for the 14 aerodrome circles with one of the events A, D, G_1, G_2, H represents a sample of the randomly presented S_1- and S_2-alternatives. For calculation of (parametric) discriminability d' the subjective responses are assumed to be drawn from independent equal variance Gaussian ($\mu_{1,2}$, σ) densities for familiarity with situations S_1 and S_2 [10]. Any discriminability difference between TWR and RTO may be quantified by corresponding coefficients $d' = \mu_1 - \mu_2 = z(H) - z(FA)$, and subjective decision bias (criterion) $c = 0.5(z(H) + z(FA))$, with z() = z-score as calculated from the inverse cumulative densities. This SDT-analysis together with Bayes inference on risk of false decision was provided in [3] for the events A, D, G1, H. In section 4.2 we will confirm these preliminary results with an additional analysis using the non-parametric discriminability index A [11] (independent of Gaussian assumption).

Table 1 lists the measured hit and false alarm rates (± standard deviations derived from binomial distributions) for the five events to be analysed with respect to A. In addition to H and FA, M = 1-H is required for calculating the total number of errors to be compared with a formal error model in section 4.3

Table 1. Measured hit and false alarm rates (H = p(y|S1), FA = p(y|S2), ± stddev from Binomial distribution according to [10]) for five events and two conditions (TWR, RTO-CWP) with a) non-answers excluded and b) non-answers added to error rates FA and M. Ta = available decision time, Tr required average decision time with stderror of mean / seconds.

Event with Alternatives S_1 / S_2 (Ta/s)	Tr / s ± stderr	CWP	a) Non-answers excluded		b) Non-answers included	
			$p(y\|S_1)$	$p(y\|S_2)$	$p(y\|S_1)$	$p(y\|S_2)$
A: bank angle var.: y / n (20)	13.8±1.7	TWR	0.92±.04	0.08±.04	0.81±.06	0.20±.05
	14.0±1.1	RTO	0.93±.05	0.11±.05	0.60±.07	0.39±.07
D: Altitude var.: y / n (20)	8.8 ±1.4	TWR	0.80±.06	0.03±.03	0.77±.06	0.12±.06
	12.4±1.5	RTO	0.73±.07	0.03±.03	0.70±.07	0.06±.04
G1: lights off: y / n (180)	27.0±6.6	TWR	0.94±.04	0.25±.07	0.94±.04	0.28±.07
	95.4±7.4	RTO	0.92±.06	0.63±.08	0.65±.08	0.72±.07
G2: Glidepath y/n (90)	21.6±6.4	TWR	0.90±.04	0.32±.07	0.88±.05	0.33±.07
	34.2±8.1	RTO	0.92±.04	0.22±.06	0.88±.05	0.22±.06
H: gear-down: y / n (10)	8.1±0.9	TWR	0.98±.02	0.06±.04	0.91±.04	0.22±.06
	9.2±0.5	RTO	0.98±.02	0.07±.05	0.77±.06	0.37±.08

Comparing the measured hit and false alarm rates for all five events under TWR and RTO conditions with non-answers not considered (optimistic case a): left two data columns), the RTO-CWP exhibits no significant difference as compared to the TWR-CWP. If however, the non-answers are interpreted as erroneous responses and correspondingly attributed to rates FA and M (pessimistic case b): right two data columns), significant differences TWR vs. RTO are obtained (smaller H(RTO), larger FA(RTO)) for event/task A (bank angle variation?), H (gear down?), G1 (lights off?), whereas for event/tasks D and G2 responses again exhibit no significant difference. The latter two tasks reflect the fact that altitude information could be read directly from the radar display and operators were free to select their appropriate information source. An extremely high FA difference TWR vs. RTO is observed for both case a) and b) for the "lights-off" event which is reflected also in a large difference of decision distance (correlated with response time).

4 Data Analysis and Discussion

4.1 Technical Limitations

Technical parameters of the reconstructed far view with videopanorama and PTZ [1][2] leads to predictions concerning performance differences under the two conditions TWR and RTO-CWP. The measured performance also depends on the usage of the different available information sources, in particular videopanorama, PTZ, and approach radar. The visibility limitations of the videopanorama are quantified by the modulation transfer characteristic (MTF), with the digital (pixel) camera resolution providing the basic limit (Nyquist criterion) for detectable objects and maneuvers: angular resolution was estimated as $\delta\alpha \approx 2$ arc min $\approx$ 1/30° $\approx$ 0.6 m object size / km distance per pixel under maximum visibility and contrast (about half as good as the human eye (1 arcmin)). Reduced contrast of course reduces the discriminability according to the MTF and the question arises how the discriminability difference TWR vs. RTO-CWP is affected. The gear-down situation at positions H1- H3 with wheel diameter 0.65 m, e.g. can certainly not be detected before the wheel occupies, say, 4 pixels which for the 40" display (0.55 mm pixel size) means a viewing angle of ca 1 mm/2 m $\approx$ 0. 5 mrad corresponding to the visual resolution of the eye (1 arcmin) under optimum contrast. This estimate results in a panorama based gear-down detectability distance of < 500 m. It means that under RTO conditions this task requires usage of PTZ in any case for enabeling a decision. The same argument is valid for the detection of bank angle changes at position A following the overflight of the runway because it requires optical resolution of the A/C-wings. The "lights-off?"-decision (G1) has a somewhat different character because in situation S_1 (lights-off, answer "yes" = hit) observers usually wait until they actually detect the A/C whereas situation S_2 can be recognized at a larger A/C distance due to the higher contrast ratio of landing-light-on/background luminance.

4.2 Discriminability of Aircraft Maneuvers during Aerodrome Circling

Based on the extended set of data as compared to [3], the focus here is on quantification of the discriminability deficit of the RTO-CWP by means of the nonparametric sensitivity index A with corrected algorithms [11], and the derivation of initial evidence for the IP/TP hypothesis [4] as formal framework for explaining the measured performance decrease. In [3] the (H, FA)-data of table 1 (without G2) were analysed using parametric discriminability d' and Bayes inference (see section 3). With decisions based on visual observation using videopanorama and PTZ, both SDT and Bayes analysis showed consistently a significantly reduced discriminability for the three maneuvers A, G1, H, but not for altitude change D where radar provided the required information. Due to the d' dependency on Gaussian distribution parameters we test here the reliability of the preliminary results with the nonparametric discriminability parameter A [11] which is calculated directly from H, FA. A is the average area under the minimum and maximum area proper ROC-isosensitivity curves (constant d', [3][10]) and varies between 0.5 (d' = 0) and 1 (lim d'-> ∞) . Figure 4 (right) depicts for analysis of case b) the A-values of the five tasks at A, D, G1, G2, H, for the two conditions TWR-CWP, RTO-CWP. Fig.4(left) shows one example of (A, b)-parametrized isopleths determined by the two TWR and RTO-CWP datapoints.

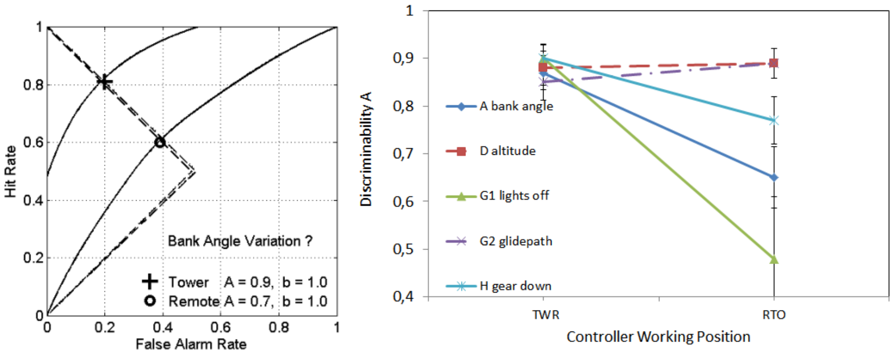

Fig. 4. Left: Isosensitivity curves (TWR, RTO) for maneuver A (solid lines, case b)-A-isopleths) and decision bias (dashed, b-isopleths). Right: A as calculated according to [11] from H and FA rates in table 1 for case b): non-answers := false answers. D and G2 (dash-dotted lines connecting TWR – RTO data): decisions about altitude (variations). A, G1, H = visual-only information (solid lines). Error bars = std. dev. based on binomial distribution [10].

The example (A, b)-isopleths for maneuver A (Fig. 4, left) shows zero decision bias (b = 1), however a significant discriminability decrease for RTO-CWP (circle = data average; minimum A-isopleth = 0.5 = positive diagonal). In agreement with table 1 and the d' results in [3], discriminability indices A in Fig.4(right) exhibit no significant difference between TWR and RTO-CWP conditions for events D, G2 (event sub-set with altitude stimulus; altitude information provided by radar), whereas the A-decrease for the visual-only subset {A, G1, H} is evident. Moreover even a reduction of the number of erroneous decisions by attributing a 50% chance to non-answers to be correct instead of assuming 100% wrong answers) leaves the RTO-performance decrease for visual-only tasks significant. The drop to chance level of RTO-CWP discriminability for case G1 is attributed to the RTO-resolution and contrast deficit which prohibits recognition of A/C even with lights on for short response times Tr: when participants at RTO-CWP after task initialization had waited some 10 s or so without recognizing landing lights they often simply guessed lights to be off or gave no answer, contributing to FA-errors.

4.3 The Information Processing / Time Pressure Hypothesis

In order to determine appropriate solutions for rising the RTO-CWP performance to at least the level of the TWR-CWP we have to find explanations for the measured discriminability deficits. The RTO-CWP performance for decision making using visual information only should be at least as good as that one based on radar used for the altitude related decisions (Fig. 4) so that users can be certain that replacement of the out-of-windows view has a potential of even improving their work condition.

A (algorithmically) simple theoretical model with some potential for explaining observed performance differences quantified in terms of decision-error probability, is based on the perceptual control/information processing theory (PCT/IP) of Hendy et al. [4]. Because our experiment was not initially designed for an application of this theory we can only expect a first impression on the relevance of the corresponding assumptions. The core idea is to formalize the information processed (as part of the total information required for a correct answer: Br / bits) as function of time pressure TP. TP is the ratio of required time Tr (to acquire Br) and the available time Ta: TP = Tr/Ta. Assuming constant cognitive processing rate (channel capacity C: Tr = Br/C) the rate of information processing demanded RID is related to TP via TP = RID/C, with RID = Br/Ta. Hendy et al. [4] derived simple algorithms for modeling dependent variables like operator workload (OWL), success ratio, and number of errors as function of TP. For the latter they suggested an exponential dependency for the increase of decision errors with TP, where TP increases linearly with the number N of objects to be analysed (in our case N = 1): TP = t_0(1 + b1 N)/Ta, and t_0 = minimal decision time for N = 0. For error probabilities we modify Hendy's algorithm in order to use our maximum error probability $p_{err} = 0.5 = p_{max}$ (guessing, no information available) as boundary condition. Keeping the original assumption that errors start to grow exponentially with TP but then level off at p_{max} we arrive at a logistic function with threshold and sensitivity parameters as one possible model:

$$p_{err} = 0.5\left(1 + exp\left\{-\left(\frac{TP-\mu}{\beta}\right)\right\}\right)^{-1} \quad (1)$$

μ ($0 \le \mu \le 1$) models the threshold where the observer starts shedding most information due to increasing workload (stress due to TP increase). It fulfills the conditions that lim(TP >> μ) $p_{err} \rightarrow 0.5$ and lim(TP → 0) $p_{err} \rightarrow 0$. The latter condition is fulfilled as long as $\mu/\beta >> 1$, i.e. steep slope (= error sensitivity $dp_{err}/dTP = 1/2\beta$ at TP = μ and/or large threshold). Figure 5 shows the results of nonlinear fitting of the respective two data points p_{err}(TP) at TP(Tr(TWR)), TP(Tr(RTO)) with the two boundary conditions (p_{err}(TP → 0, TP → ∞) using model-equation (1) for the three visual-only tasks. For characterising the experimental results in terms of (μ, β) we have to use the total number of errors for the full set (n(S1)+n(S2)) of trials per subject instead of the conditional probabilities, misses and false alarm rates M=1-H, FA: p_{err} = (n1 M + n2 FA)/(n1 + n2) as used for the discriminability calculation.

The results indicate the principal applicability of the logistic error model because all three cases yield reasonable threshold (μ < TP = 1) and error sensitivity parameters $1/\beta$. The RTO-performance deficit always seems to correlate with some kind of time pressure. According to IP-theory decision errors should increase significantly due to increasing stress when Tr approaches Ta and to shedding of information when Tr > Ta (Tr/Ta > 1). This is reflected by our results only for event H (gear down) with the shortest Ta = 10 s. Variation of threshold μ with event(stimulus) can be explained by the fact that the three specific events provide quite different stimulus conditions for the decision making as described in section 3.The fact that only for the gear-down task an approximately exponential increase of errors at TP ≈ 1 is observed according to [4] with $\mu \approx 1$ whereas a sensitive threshold behavior at lower μ is suggested for tasks A, G1, indicates at least one more performance limiting factor besides time pressure, such as PTZ-camera contrast/resolution and operator training. For lights-off decision the RTO-HMI contrast deficit should play a major role: the average response appears completely at random. Nevertheless also in this case a long waiting time after beginning to gather visual evidence might lead to increasing stress due to uncertainty.

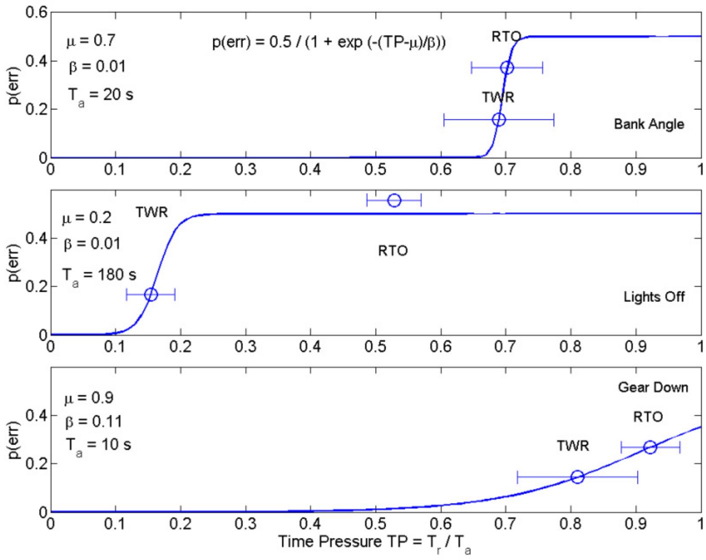

Fig. 5. Decision error probabilities for TWR and RTO-CWP vs. time pressure TP (±stderr of mean, n = n(error)+n(correct) = 80 …100) for tasks where visual / PTZ-information was used for decision making. Standard errors of p(error) are smaller than the circles of data points. Logistic error model (equ. (1)) derived from IP/TP-theory [4] for fitting p_{err}(TP).

5 Conclusion

The present analysis of two-alternative decision making with safety related aircraft maneuvers confirms the previously reported [3] explanation of an observed discrepancy in the percentage correct analysis (p_c, neglecting non-answers) [9] of the corresponding observation data, as compared to the subjective success criteria. The perceived safety was rated as insufficient by participants which agrees with the objective data of the present analysis and [3]. Neglecting non-decisions during simultaneous decision making at TWR- and RTO-CWP yields mostly no significant difference of discriminability (i.e. suggesting sufficient RTO performance) whereas the interpretation of non-decisions as false responses (misses or false alarms) leads to significant error increase under RTO as compared to TWR conditions and correspondingly reduced A and d'. The results indicate a usability deficit of the RTO-HMI (videopanorama and PTZ) due to time pressure as one possible reason. Data analysis with a modified version of the Hendy et al. information processing / time pressure theory (IP/TP) [4] indicates additional origins of performance decrease due to threshold behavior of decision errors significantly below the TP = 1 value. It is expected that increased automation (e.g. automatic PTZ-object tracking and data fusion with approach radar) will increase usability, and in combination with improved operator training could solve the performance problem. However further experiments are required for clarifying the role of time pressure and validating the effect of a higher level of automation system. They are preferably realized as human-in-the loop simulations with appropriate design for time pressure variation, and forced choice tasks for avoiding non-answers. Because of the significant effort required for the HITL-experiments and field tests, the initial results of the IP/TP-model suggest as intermediate step computer simulations for preparing corresponding HITL- and field experiments. For this purpose the commercial tool IPME (Integrated Performance Modeling Environment [12]) appears useful which integrates the PCT/IP-based approach together with a resource based theory so that by means of simulations it would allow for further clarification of the influence of different performance shaping functions.

Acknowledgement. We are indepted to DFS personnel N. Becker, T. Heeb, P. Distelkamp, and S. Axt for cooperation during preparation of the experiment. Many thanks are due to C. Möhlenbrink and A. Papenfuß for support and performing online interviews during the exercises. Markus Schmidt and Michael Rudolph were responsible for the technical setup and RTO-software of the experiment. A. Grüttemann was responsible for the onboard data acquisition and served as flight engineer. We acknowledge the support of the DLR flight experiments department and in particular cooperation with pilots G. Mitscher and P. Bergmann.

References

[1] Schmidt, M., Rudolph, M., Werther, B., Möhlenbrink, C., Fürstenau, N.: Development of an Augmented Vision Videopanorama Human-Machine Interface for Remote Airport Tower Operation. In: Smith, M.J., Salvendy, G. (eds.) Human Interface, Part II, HCII 2007. LNCS, vol. 4558, pp. 1119–1128. Springer, Heidelberg (2007)

[2] Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., Papenfuß, A., Kaltenhäuser, S.: Steps towards the Virtual Tower: Remote Airport Traffic Control Center (RAiCe). In: Proc. EIWAC 2009, ENRI Int. Workshop on ATM & CNS, Tokyo, March 5-6, pp. 67–76 (2009)

[3] Fürstenau, N., Friedrich, M., Mittendorf, M., Schmidt, M., Rudolph, M.: Discriminability of Flight Maneuvers and Risk of False Decisions Derived from Dual Choice Decision Errors in a Videopanorama-Based Remote Tower Work Position. In: Harris, D. (ed.) EPCE/HCII 2013, Part II. LNCS (LNAI), vol. 8020, pp. 105–114. Springer, Heidelberg (2013)

[4] Hendy, K.C., Jianquiao, L., Milgram, P.: Combining Time and Intensity Effects in Assessing Operator Information-Processing Load. Human Factors 39(1), 30–47 (1997)

[5] Hannon, D., Lee, J., Geyer, T.M., Sheridan, T., Francis, M., Woods, S., Malonson, M.: Feasibility evaluation of a staffed virtual tower. The Journal of Air Traffic Control, 27–39 (Winter 2008)

[6] Ellis, S.R., Liston, D.B.: Visual features involving motion seen from airport control towers. In: Proc. 11th IFAC/-IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes, France, September 31-October 3 (2010)

[7] Van Schaik, F.J., Lindqvist, G., Roessingh, H.J.M.: Assessment of visual cues by tower controllers. In: Proc. 11th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Valenciennes, France, August 31-September 3 (2010)

[8] Watson, A.B., Ramirez, C.V., Salud, E.: Predicting visibility of aircraft. PLoS One 4(5), e5594 (Published online may 20, 2009), doi:10.1371/journal.pone.0005594

[9] Friedrich, M., Möhlenbrink, C.: Which data provide the best insight? A field trial for validating a remote tower operation concept. In: Proc. 10th USA/Europe Air Traffic Management Research and Development Seminar (ATM 2013), June 10-13 (2013)

[10] MacMillan, N.A., Creelman, C.D.: Detection Theory. Psychology Press, Taylor and Francis, New York (2005)

[11] Zhang, J., Mueller, S.T.: A note on ROC analysis and non-parametric estimate of sensitivity. Psychometrica 70(1), 203–212 (2005)

[12] Fowles-Winkler, A.M.: Modelling with the Integrated Performance Modelling Environment (IPME). In: Proc. 15th European Simulation Symposium, SCS European Council (2003) ISBN 3-936150-29-X