Simulating radio communication complexity in regional airspace design

Neale L. Fulton*, Mark Westcott*

*CSIRO Mathematical and Information Sciences (CMIS), Canberra, ACT 2601, Australia
and

Stephen Emery^

^School of Civil and Environmental Engineering, University of Witwatersrand, Johannesburg, South Africa

In the early 20th century, pilots flew without any inter-aircraft communication. As traffic movements grew, the earth's atmosphere was partitioned into relatively small volumes called sectors. A low-level "area" sector reaching to ground level covered a wide area and typically enclosed many different airports. For the purpose of this paper, a particular radio frequency was assigned to the sector for the specific purpose of enabling communication between all the many aircraft at the different airports and the many enroute aircraft and, for the further purpose, of enabling aircraft-to-ground communication. As air traffic management (ATM) systems evolved and traffic flows increased, new and smaller volumes around airports (called Mandatory Broadcast Zones (MBZ) or Common Traffic Advisory Frequencies (CTAF)) were introduced with additional frequencies. However, as a consequence, communication transactions and protocols became more complex and the larger area sectors became isolated in frequency from the smaller volumes surrounding the airports.

In this increasingly complex environment pilots still must be able to communicate with each other if their aircraft come into proximity. This requirement applies regardless of the relative position and aspect of the aircraft pair within either the same or adjacent sectors. Pilot-to-pilot communication must be achieved with a high dependability (a reliability and safety concept) of the various communication links operating between the aircraft at the time of proximity. In general, a communication transaction may require a number of physical transmissions to implement the radio procedure protocols. The time taken to complete a transaction will vary based on the protocol requirements but this finite time introduces both engineering and operational complexity into the design. A broadcast is a particular type of transaction used to promulgate an aircraft's position and immediate intentions. The broadcast consists of only one physical transmission and is the focus of this study.

This paper is concerned with assessing the physical feasibility of inter-pilot communication when their aircraft are in, or near, radio frequency structures such as MBZ or CTAF. It examines how the structures might affect the dependability of communication between the aircraft. The distinctive feature of the structures studied is that communication occurs on different frequencies at different points in airspace. This means that pilots in relatively close spatial proximity might not be operating on a common frequency. When combined with relatively long transaction times, this has the potential to fatally inhibit timely exchange of information critical to successful avoidance of a midair collision. The design question that arises, regardless of the precise geometrical description of such structures, is:

> *What impact do radio frequency structures have on the dependability of pilot-to-pilot communication that must exist between the proximate aircraft in order to manage that proximity?*

We show that both normal operational modes and failure modes arise for the MBZ/CTAF structures. The modes discussed are similar to failure "modes" of operation identified in Flight Management Systems designs. These failure modes (e.g., mode confusion within the pilot-machine interface) are discussed in the aerospace literature.

The paper uses a simple but revealing model of aircraft operation within a multiple radio frequency structure to study the operational modes. The model is not intended to be complete or exhaustive; its role is to demonstrate design principles and processes that should be considered in order to achieve required levels of system design confidence. One important conclusion is that circumstances in which problems can arise are not easily predictable during in-flight operations. This means that operational experience is not necessarily a good basis on which to predicate the extrapolation of system design behaviour, as aircraft might often be operated close to a failure mode without the pilots realising it and so they may erroneously conclude these modes do not exist. The model allows an exhaustive description of the failure modes once parameters such as aircraft speeds and headings, radio frequency structures and communication transaction lengths are specified. To show how the failure modes are influenced by these parameters, the paper uses a novel form of nested plot for high-dimensional data that was developed for similar displays in an independent study for a large commercial contract.

***Keywords*:** aircraft proximity, flightpath arrangement, partition, bipartite graph, cluster.

1. Introduction

Historically, one radio frequency serviced many airports in a specified region. More recently (1991 - 2009) circular cylindrical airspace structures, located at an airport, have been introduced in Australia. The rationale presented is to contain communication associated with circuit traffic to special frequencies, removing the need to service airport communication loads from the area frequencies. Two similar structures have been introduced; CTAF and MBZ (AIP, 2001), as illustrated in Fig1.

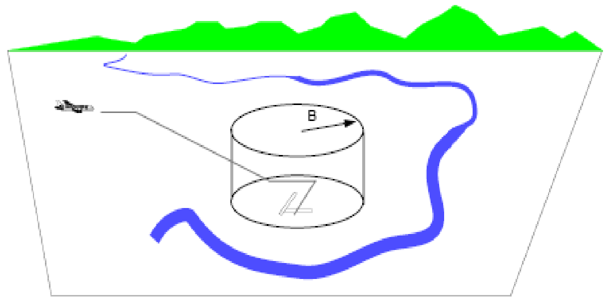

Fig1: CTAF/MBZ structure

For our purposes, the only difference between the CTAF and the MBZ is the radius, B, of the cylinder. The radius is typically 5 NM for the CTAF, but may range from 5 NM to circa 30 NM for the MBZ. Radio procedures prescribe the use of an area frequency to broadcast intentions when operating outside the cylinder and on a second frequency when entering or operating within the cylinder. To avoid mid-air collisions, pilots must be to able to inform other aircraft of their present position, speed and track. While the intent of these structures is clear it is also clear that communication between aircraft can degenerate when aircraft operate near to or have to transition the MBZ/CTAF boundary. This arises because two aircraft in close proximity can have different frequencies selected on their radios.

A fundamental design requirement is for pilots of proximate aircraft to be able to achieve a prescribed alert time. Fulton (2002) and Fulton, Baird and Smith (2003, 2002) have investigated these times when aircraft cross radio frequency boundaries. For the simplest case, the absence of a radio frequency boundary, when aircraft operate on the one area frequency, most aircraft can achieve a five minute alert time if they report inbound at 30 NM. Only the fastest aircraft with the highest relative closing speed (250 - 500 Kts) fail to meet this criterion achieving instead a three minute alert when relative closing speed combinations are in the range 250 - 500 Kts. When a radio frequency boundary is introduced, these alert times can no longer be guaranteed. Pilots then need to synchronise frequencies in order to be able to communicate (note that the airspace design must work for aircraft equipped with only one radio). In this situation it is essential that an alternative mode of communication be used to back up the communication links while the pilots make the transition from one radio frequency to the next. If an independent and alternative mode of communication is not available then the high dependability (see Villemeur, 1992) of operation cannot be assured.

This study and others referenced, show that removal of ground-based traffic information, or tower services means that pilots may not have sufficient time to communicate relevant information using long transaction time communications such as direct Pilot-to-Pilot Voice Link (PPVL). The paper does not advocate the removal or making voice communication obsolete but rather emphasises the care needed in the design of communication systems. For example, failure can arise because critical calls may be blocked; calls may not be scheduled in the correct priority order; the channels may saturate during times of peak demand. A poor system design becomes open to various failure modes.

2. A hybrid design model for an MBZ/CTAF

In assessing the performance of an MBZ/CTAF structure it is critical to consider both the kinematics of aircraft proximity and the logical communication process needed to control that proximity, and the interaction between them. The main points we emphasise are:

- There are certain combinations of kinematic and logical conditions where MBZ/CTAF structures will fail.
- ICAO (2001) Annex 11 requires continuity of service.
- The MBZ/CTAF procedures when operated under self-separation provide no warning as to when failure will occur.
- If self-separation (segregation) functions are augmented by separation functions then failure instances can be reduced. However, this is no guarantee, as demonstrated by Sioux Lookout accident (TSB-Canada, 1995).
- Pilots may use the procedures many times without failure only to find that, on the next operation, a small deviation in procedures or timing takes the system to a failure mode (see Degani (1996) on flight management systems).

The model

At time zero, an outbound aircraft, O, takes off from an airport. There is another (inbound) aircraft, I, in the area on a collision course with O. Each is initially unaware of the presence of the other. So they will collide unless they communicate in time to take appropriate evasive action. Note that I could be planning to land at the airport or just be in transit in the vicinity of the airport. We make the following assumptions throughout.

Kinematic assumptions

K1. Each aircraft, in each scenario, has a prescribed constant resultant velocity vector, $\mathbf{V}_O$ and $\mathbf{V}_I$, respectively.

K2. The initial position of I is such that O and I will collide if each aircraft achieves its resultant velocity vector.

Communication assumptions

C1. There are two communication frequencies used; an *inner frequency* f_1 in the neighbourhood of the airport, and an *outer (area) frequency* f_2 elsewhere.

C2. Aircraft on different frequencies cannot communicate.

C3. There can be only one transmission at a time on a frequency; all other transmissions on the frequency are *blocked*.

C4. O broadcasts on f_1 as it takes off. It then broadcasts on f_2 as it exits a cylindrical region C_b of radius *b (<B).*

C5. I broadcasts on f_1 as it enters a cylindrical region C_B of radius *B*. If it later exits C_b it will broadcast again on f_2.

C6. The transmission lengths for each aircraft, u_O and u_I, are constant but possibly different.

C7. If the aircraft communicate successfully they will avoid a collision. Successful communication means that a *complete* transmission is made by one aircraft and received by the other aircraft.

Other assumptions

A1. O and I can manoeuvre instantly to avoid a collision, so pilot and aircraft control reaction times are ignored.

A2. B > b (this is to avoid extra complications, for purposes of exposition)

A3. I is not transmitting on f_1 at time 0, so the initial transmission by O is never blocked.

Assumption A3 will sometimes not be used. These assumptions are clearly simplistic. However the results that come from them show that there is a considerable variety and complexity of possible interactions between O and I even in this case, which is the main point of the paper.

The plane defined by $\mathbf{V}_O$ and $\mathbf{V}_I$ is called the *engagement* plane, and all planar figures are drawn in this plane. So we can define a Cartesian coordinate system in this plane, with origin at the airport, and without loss of generality we assume O flies along the *x*-axis; thus $\mathbf{V}_O = V_O(1,0)$, where V_O is the speed of O. We write V_I for the speed of I, and γ for the *speed ratio*; thus $\gamma = V_I / V_O$.

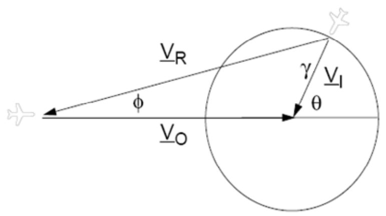

Fig2: Velocity space for proximate aircraft

Assumption K2 implies that the *relative velocity vector* $\mathbf{V}_R = \mathbf{V}_I - \mathbf{V}_O$, which also lies in the engagement plane, must go through the origin. This is a partial explanation for the complexity of the results; the kinematics are most naturally analysed in relative velocity space while the communications rules relate to absolute space. The situation is shown in Fig2, where θ is the track intercept angle (it is 180 – the relative heading of I) and ϕ is the polar angle of $\mathbf{V}_R$. Another way to view K2 is that the position of I at $t = 0$ is on a line through the origin at angle ϕ. Its (polar) distance from the origin is written as either $\xi(0)$ or ξ_0. The *time to collision,* t_c, is then given by $t_c = \|\mathbf{V}_R\| \xi(0)$.

3. Operational modes - success and failure

Because we have assumed that the two aircraft can react instantly to avoid a collision once they have communicated fully, it follows that *a collision can only occur during a transmission by one or other aircraft*. This means that the problem naturally partitions into a number of *communication regimes*. Provided the track of I does not exit C_b before meeting the track of O, these are defined as follows (assuming A3). Here, t_I is the time when I enters C_B and t_O is the time when O exits C_b.

Regime	Description	Definition
Regime 0	I enters C_B before time 0, so it will hear the initial transmission by O	$t_I < 0$
Regime 1	I enters C_B while O is still transmitting. It tries to transmit (on f_1) but is blocked.	$0 \le t_I < u_O$
Regime 2	I enters C_B after time u_O but before O exits C_b, and transmits on f_1.	$u_O \le t_I < t_O$
Regime 3	I is still outside C_B when O exits C_b and transmits on f_2.	$t_I \ge t_O$

If we do not use A3, Regime 0 becomes two regimes, Regimes 0a and 0b, say. These are defined as follows.

Regime 0a	I enters C_B before time $-u_I$, so finishes transmitting (on f_1) before O takes off	$t_I < -u_I$
Regime 0b	I enters C_B at time between $-u_I$ and 0, so its transmission blocks that of O at takeoff	$-u_I < t_I < 0$

Each regime is closely associated with a *communication failure mode*, or *mode* for short, as follows.

Mode	**Description**	**Definition**
Mode 0	Collision during initial transmission by O.	$t_c \le u_O$
Mode1	Collision during transmission by I that was initially blocked by O's transmission; i.e. during Regime 1.	$u_O < t_c \le u_O + u_I$
Mode 2	Collision during transmission by I at entry to C_B; i.e. during Regime 2	$t_c \le t_I + u_I$
Mode 3	Collision during transmission by O at exit from C_b; i.e. during Regime 3	$t_I \ge t_O$

Again, if we do not use A3 then Mode 0 becomes two modes.

In principle, Mode 0 can occur during either of Regimes 0 or 1. It cannot occur during Regimes 2 or 3, since in either case $u_O \le t_I < t_c$. In practice, it is likely to occur only during Regime 0. The other three Modes are specifically associated with a single Regime. In practice, all regimes and modes are likely to be realisable; in principle, some may be void. For example, if $v_O u_O > B$ then avoidance of a Mode 0 failure puts O outside C_B and hence beyond Regime 1.

The requirement that "the track of I does not exit C_b before meeting the track of O" always holds for θ acute, and in some cases for θ obtuse. If it does not hold, we have to include the extra possibility of interacting communications when both O and I are exiting C_b.

4. Analysis of the head-on case (with A3)

To illustrate the concepts in the simplest mathematical setting, which still illustrates most of the possible complexities, we analyse the case where the two aircraft are approaching head-on (θ = 0). Clearly $\phi = 0$. Further, it is easy to see that $\xi_0 = (v_O + v_I)t_c$ and $t_I = (\xi_0 - B)/v_I$ .Assume that $u_O + u_I < t_O$; this ensures that all four Regimes can occur. Then:

Regime 0 occurs when $\xi_0 \le B$;
Regime 1 occurs when $B < \xi_0 \le B + u_O v_I$;
Regime 2 occurs when $B + u_O v_I < \xi_0 \le B + t_O v_I$;
Regime 3 occurs when $B + t_O v_I < \xi_0$.

Now we look at the associated failure Modes. We want to express the condition for each Mode in terms of ξ_0. This is straightforward for all Modes except Mode 2, because of the simple connection between the defining t_c and ξ_0 given above. Mode 2 is more complicated in general because the Mode definition involves t_I, which is itself a function of ξ_0. However, we must be careful with cases where certain Modes become irrelevant.

Mode 0 occurs when $\xi_0 \le (v_O + v_I)u_O$.

Mode 1 occurs when $u_O < t_c \le u_O + u_I$, or equivalently $(v_O + v_I)u_O < \xi_0 \le (v_O + v_I)(u_O + u_I)$ during Regime 1, *provided* this does not intersect with Mode 0; that is, provided $(v_O + v_I)u_O \le B$. Otherwise, we have two possibilities:

a. $u_O v_O < B \le (v_O + v_I)u_O$. In this case the lower boundary of the Regime 1 region is $(v_O + v_I)u_O$ rather than *B*.
b. $B \le u_O v_O$. In this case Regime 1 is totally subsumed in Mode 0 and hence there can be no Mode 1 failures.

Combining these facts gives $\max\{B, u_O(v_O + v_I)\} < \xi_0 < \min\{B + u_O v_I, (u_O + u_I)(v_O + v_I)\}$ as the characterization of Mode 1. From b, this inequality is void if $B \le u_O v_O$. In this case, there can be no Mode 1 failures.

Mode 2 requires $t_c \le t_I + u_I$. Substituting from earlier equations for t_c and t_I gives $\xi_0/(v_O + v_I) \le (\xi_0 - B)/v_I + u_I$, whence $\xi_0 \ge (1+\gamma)(B - v_I u_I)$. So the full condition for Mode 2 is $(1+\gamma)(B - v_I u_I) \le \xi_0 \le B + v_I t_O$, since Mode 2 can only occur during Regime 2. Of course, this inequality may be impossible to satisfy for a particular choice of *B*, *b*, v_O, v_I and u_I, in which case there is no initial position for the inbound aircraft that will result in a Mode 2 failure.

Mode 3 occurs when $B + t_O v_I < \xi_0 \le (v_O + v_I)(t_O + u_O)$.

There are various plots that can illustrate these results. In Fig3 we plot the track of O and a range of tracks possible for I, specifically the Regime and Mode boundaries. The coordinates are time and the radial distance from the origin.

The parameter values are $v_O = 120$, $v_I = 180$, $u_O = 1$, $u_I = 1.5$, $\theta = 0$ The dotted black lines sloping downwards are the Regime boundaries (with 0a and 0b); the solid coloured lines are Mode boundaries. The solid step line is the communication boundary. The fate of a particular I, that is a particular downward sloping line defined by choice of ξ_0 (the *y*-intercept), depends on whether it meets the communication boundary before or after it meets the track of O. If before, then communication occurs and the potential collision is avoided; if after, the collision occurs.

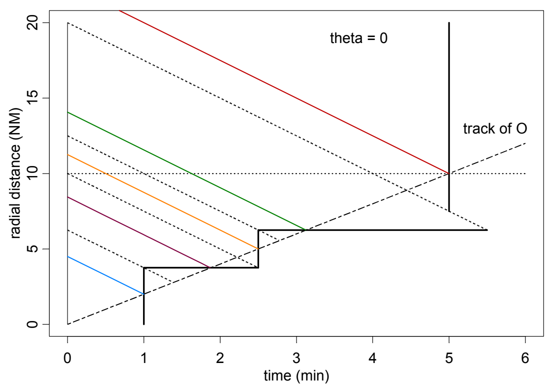

Fig3. Plot of radial distance against time for O and I, showing the Regime and Mode boundaries

Fig4 shows basically the same situation (though with A3) but in a plot of 'initial position' against speed ratio γ. Since this plot has been standardised by dividing by v_O, the 'initial position' is actually ξ_0 / v_O, a time. Here, the safe region is to the left of the solid line while collisions occur to the right. The colours of the Regime boundaries match those in Fig3.

The dashed vertical line in Fig4 is at $\gamma = 1.25$, which is equivalent to the choice of speeds in Fig3. The switching between safety and collision as ξ_0 varies along the dotted line summarises Fig.3. Fig4 shows the outcomes for a range of γ but it lacks the dynamic element of Fig3. Both plots are useful summaries.

When θ is obtuse, the green boundary can curve back on itself, so can be cut twice by a vertical line. This shows the added complexity that can occur for such θ.

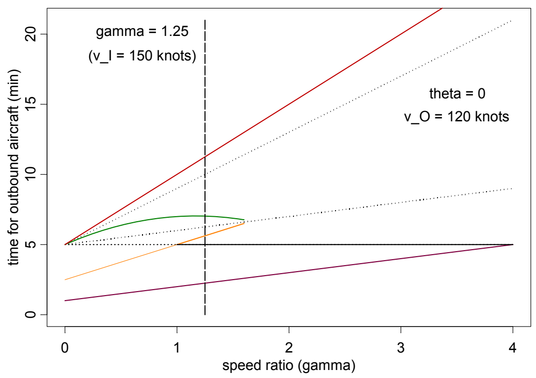

Fig4. Plot of speed ratio against 'initial position' (see text) , showing the Mode boundaries. The collision region is to the right of the solid line.

The third plot shows the results for a large number of cases (sets of parameter values) simultaneously. It uses a novel method for plotting high-dimensional data, which was developed for another project by the CMIS authors with David Gates (CSIRO Mathematical and Information Sciences). The MBZ/CTAF design space in our model has eight parameters: $b, B, u_O, u_I, v_O, v_I, \theta, \xi(0)$. Our plot, called a *nested plot,* shows results for cases in which six of the eight parameters vary; for each individual plot v_O and u_O are fixed. There are then three levels of nesting, as shown in Fig5. At the top level there is a 3x3 array for the values of B and u_I. Within each of these 9 rectangles is a 4x2 array for the values of θ and *b*. Finally, within each of these 8 rectangles is a 30x10 array for the values of $\xi(0)$ and v_I. By this means, results for a total of 21,600 cases can be shown on a single plot, colour-coded by outcome; here, the outcome is successful communication or collision, with the latter further coded by the same colours as in Figs 3 and 4 to show the failure mode. Note that each inner rectangle in the nested plot is essentially a discretization of Fig4 (with the axes reversed). The number of levels of nesting, and of cases included, is partly determined by the plotting resolution, to ensure individual cases, or pixels, can be clearly distinguished.

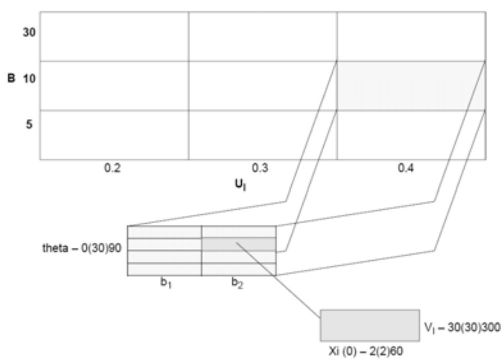

Fig5. The template of the nested plot

Fig4 shows the results of the 21,600 cases with $v_O = 240$ kts and $u_O = 0.6$ min. From the plot we can readily see trends in the failure patterns as the design variables change. For example: increasing *B* has a significant effect on the pattern whereas increasing u_I has very little effect; the results for $\theta = 0, 30$ and 60 are fairly similar but perceptibly different from those with $\theta = 90$; Mode 1 failures are rare except at the lowest value of *B*. Further, the results for $\theta = 90$ often show just a few failure cases somewhat isolated from each other. This is another instance of the points we wish to make in the paper:

- failures can be unlikely, which ensures that experience might not be a good teacher;
- failures are hard to foresee or anticipate because the circumstances which cause them are little different from those leading to a safe outcome.

A series of such plots for different v_O and u_O gives an easily assimilated picture of the effects of variables and trend as they change. In the study for which this model was developed, we displayed the results for over 1 million cases by this means.

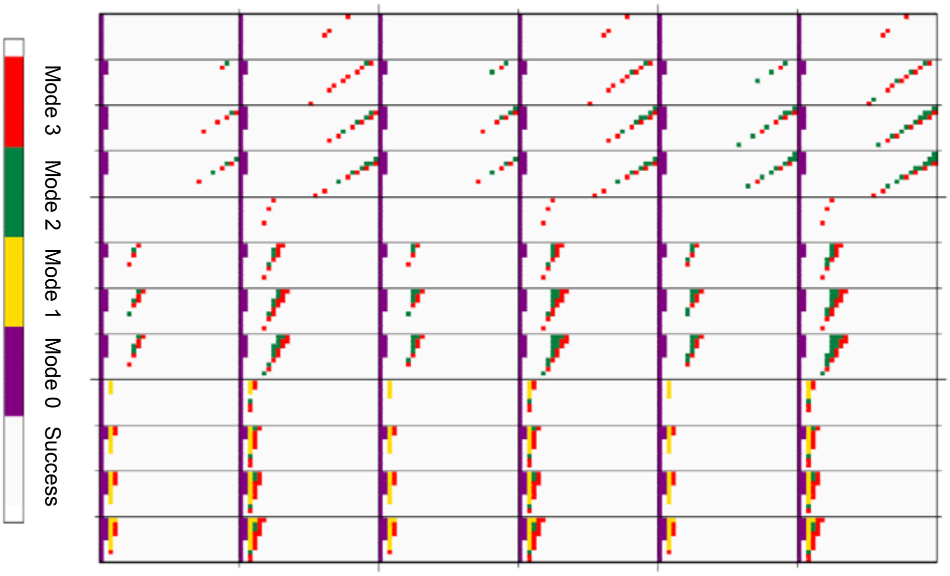

Fig6. Nested plot of outcomes for cases with $v_O = 240$ kts and $u_O = 0.6$ min. The template for this plot is given in Fig.5

5. Discussion

Direct voice radio communication between aircraft, particularly in the presence of a MBZ/CTAF structure has been investigated. Four modes of failure have been identified. These modes will, a priori, be transparent to pilots and should one of these modes be encountered the system will fail catastrophically, that is, without warning.

The present analysis is based on constant velocity vectors and therefore straight line tracks. It is a first order representation that illustrates the problems without undue detail. *More realistic kinematics would change the detail but the same qualitative conclusions would emerge.* The communication policy selected is a reasonable interpretation of current operational practice. Of course, navigation errors, weather, pilot task loads and other factors all contribute to the outcome so an analysis using this model, though extensive, represents only part of the domain of possible behaviours.

The operational consequences of our results are as follows. A pilot may operate in and out of an MBZ/CTAF many times and not experiencing failure, but then a combination of circumstances (design variables) occurs that moves the system from a safe to a failure mode. *And this can happen without the system providing any warning it will now fail.* There is a strong analogy between the type of failure experienced in airspace and that of software in computing; certain in-flight situations (threads) will always work, others will always fail.

Our results may partly explain the differing perceptions of different user groups within the aviation community. The performance of the MBZ/CTAF structure is highly sensitive to the closing speed between aircraft and the radius of the structure. So two slow gliders might claim the communication protocol always works; two high performance RPT aircraft operating at the legal maximum of 250 KTs (below 10,000 FT) might claim frequent difficulties with the protocol while two intermediate speed GA aircraft might claim that the protocol works most of the time.

The clear conclusion is that there will be situations in which unaided segregation by PPVL will at times fail, in both the MBZ and the CTAF structure. Segregation needs to be augmented by some other form of communication that, jointly with PPVL, creates a fault tolerant system. This will typically be by a ground based agency but with modern mobile communications such augmentation may be implemented by an air-to-air function such as aircraft-to-aircraft exchange of position information and pilot alerting by data-link.

Historically, the practical measurement of flightpath activity in unmanaged airspace has been a difficult problem due to the lack of systematic real-time reporting and recording of aircraft position. In addition sporadic and often low frequency of activity has meant that accumulated operational experience cannot populate the combinatorially large state-space even after decades of use. In terms of airspace design and analysis our model provides a systematic mathematical basis by which the operational state-space can be specified, populated and explored. It allows the adequacy and feasibility of various communication models to be tested and compared both by analysis and simulation.

In terms of operational test the practical difficulties of measuring in real-time, aircraft position, flightpath propagation and interaction will be largely overcome, as data-link communication between aircraft becomes an operational reality. In this situation the mathematical modelling presented will permit a more accurate dynamic forecasting of system performance and limitations. In particular, data-link reporting of position will also provide a foundational specification on which a more accurate prediction of the actual performance of still essential aircraft-to-aircraft voice communication can be based.

Acknowledgements

CSIRO wishes to thank Michael Caplehorn, Chairman of Broome International Airport, for his assistance and enthusiastic support of the research project, and his permission to use the data for publication.

References

AIP (2001), Aeronautical Information Publication, Amendment 31, Airservices Australia, Canberra, Australia, August 10.

Degani, A., (1996), Modelling Human-Machine Systems: On Modes, Error, and Patterns of Interaction, PhD Dissertation, Georgia Institute of Technology, December.

Fulton, N., L., (2002), **Regional Airspace Design: A Structured Systems Engineering Approach,** Doctoral Dissertation, University of New South Wales, (Submitted 5 July 2002), 11 December.

Fulton, N., L., Baird, J., Smith, W., (2003), **A Compromise Decision Support Problem Approach to Airspace Design,** In Proc. 10th Australian International Aerospace Congress/ 14th National Space Engineering Symposium, Brisbane, 29 July- 1 August.

Fulton, N., L., Baird, J., Smith, W., (2002), **Modelling Airspace Design from a Systems Engineering Perspective,** In Proc. 9th AIAA/ISSMO Symposium on Multi-disciplinary Analysis and Optimisation, Atlanta, 4-6 September.

ICAO (2001), Annex 11 to the Convention on International Civil Aviation, Air Traffic Services, Thirteenth Edition, July

Kubu (2003), Broome Terminal Airspace Aeronautical Study Preliminary Analysis, Kubu Australia Pty Ltd, July.

Kubu (2004), Design Aeronautical Study for Broome International Airport Terminal Airspace, Kubu Australia Pty Ltd, July.

TSB-Canada (1995), Mid-Air Collision Between Bearskin Airlines Fairchild Metro 23 C-GYYB and Air Sandy Inc. Registration PA-31 Navajo C-GYPZ - Sioux Lookout Ontario, 12 [sic] nm NW 01 May 1995. Report No. A95H0008, Transportation Safety Board of Canada.

Villemeur A., (1992), Reliability, Availability, Maintainability and Safety Assessment. Anne Cartier, Marie-Christine Lartisien (Translators), John Wiley & Sons, Chichester.