AGENDA ITEM 16.2

CONTEST SCORING SYSTEM REPORT

ACRO

Nick Buckenham

Synopsis

There were five CIVA international championships during the 2009 season, the ACRO scoring system being used to manage the contest data and provide the results at each event:

Event:	*Scorer:*
The World Air Games, Italy	Maurizio Carrannante
The World Yak-52 Aerobatic Championship, Lithuania	Vytautus Tautkevicius
The World Glider Aerobatic Championships, Czech Republic	Martin Götz
The European Advanced Aerobatic Championship, Poland	Jürgen Leukefeld
The World Aerobatic Championship, UK	Jürgen Leukefeld

In each case the scorer was able to enter the personnel, aeroplane and judging output quickly and efficiently, provide detailed information reports for contest staff where appropriate, process the results data and upload interim and final results to the CIVA Results web without delay.

A suggested change to the rules for judging flick-rolls to resolve a problem that particularly impacts the FPS Ranking Index is presented for discussion.

Issues resolved during 2009

A small number of low-level technical problems were experienced whilst using the software, arising from unusual aspects of the structure of the event – for example the need to operate with more than 20 figures per sequence for some pilots at the WAG, and the presence of some non-Latin characters in pilots names affecting the linking of online score-sheets to their tabulated scores. A change was also necessary following the 3rd programme of WGAC when the requirements of Section 6, Part 2, section 4.1.8 regarding the re-ordering of the pilot Groups for the 4th to 6th programmes were not fully met.

On each occasion the situation was resolved without impact to the event itself, either through

advice on how to work around the problem or by making suitable changes to the software itself and transmitting the new code to the event so that the scorer could quickly adopt it and continue.

FairPlay

The 'FairPlay' statistical processing system continues to provide a sound basis for evaluating pilot performances and the foundation data for post-sequence analysis of judges marks. The output from FPS appears well accepted, though while most pilots have a general idea of its purpose it seems not many know how the process achieves it. A better understanding amongst users should lead to improved confidence in the system's output, and help non-users to see the benefits FPS provides for pilots and in the area of judge development.

In the past I have tried to tackle these explanations via step-by-step explanations of the logical process, but this is too complex to be effective. The vigorous reaction by one pilot at WGAC following abortive attempts to match his raw grades and figure K-factors with his countries judge analysis and the FPS calculated results has reinforced the need for simpler grass-roots explanatory material to be available for distribution at competitions to clarify:

- Why we need FPS – judging experience and styles, bias etc..
- The logical process from judges marks to final scores.
- For pilots how to 'read' the online score-sheets, and for judges their personal analyses.

It might be optimistic to believe that this will persuade those not using FPS to adopt it, but that is less likely to happen unless some convincing steps such as this are put in place.

The Ranking Index

CIVA's use of this derived index to assess how well judges perform is now well established, and appears to provide a sound basis for separating good, indifferent and poor judging at major events. We mustn't loose sight however of the RI's *relative* view of the judges at each event: FairPlay and the computer do not know how to judge, they can only report judges who are less 'in-tune' with the majority than others. A useful analogy here is to view a perfect zero RI as white and above perhaps 40 as black, with notional interim grey stages around 15 and 25 where a judge's performance passes from acceptable to questionable, and subsequently to unacceptable …. these boundaries are not rigid and depend also on the overall size of the event and the data set, the standards of the judges, and probably the circumstances at the event itself.

However I think that in the judging of figures with flick-rolls we have an RI driven problem that we should try to settle. When assessing badly initiated flicks many judges shy away from giving a Soft Zero because they see it as a 'risky' decision that might harm their RI, and consequently are more likely to give an over-cautious 'safe' grade.

- If a flick-roll clearly fails the initial pitch / yaw criteria and the judge is confident that his colleagues will concur then an appropriate Soft Zero is likely to be given.
- However … if the initiating characteristics of the flick are borderline – for example with only a small amount of pitch and yaw – the judge may be over-concerned that an SZ verdict will become an anomaly when compared to more cautious judges observations, and compensate in some other way.

A minority SZ is very likely to fail the FPS confidence test. The judges ranking of that pilot will then be affected, and when a high-ranking pilot is involved the 'hit' to the judges RI can be significant. I know from discussions with many judges that this often leads to the tough decision being ducked and a downgrade (not in the rules!) given instead. This is a particular concern to less experienced judges and those who are new to FPS.

Section-6 6.8.24 and 25 explicitly require flicks with no or inadequate initial pitch/yaw to receive a SZ, but we quite often see minimal pitch/yaw – frequently with visible aileron application – to 'sham' the flick. I think that the problem of how to judge this could be considerably simplified if we add the option of a fixed downgrade to a flick that may technically be autorotation but does not adequately display the required characteristics at onset. If judges could conscientiously apply a rule based downgrade – for example 4 marks – to a 'shallow' flick rather than duck the issue of a SZ to protect their RI then the reduced mark would be more likely to survive FPS intact and quickly concentrate pilots minds on showing more convincing figure execution, a two-fold benefit.

Judging analysis

This analysis continues to be the primary – indeed the only practical – source of immediate feedback to all judges following completion of each sequence of flights. It is however seen as a daunting set of tabulated data, and we should aim to make it easier to read and interpret. During the year I have added an additional set of pages to the existing Judge Individual Analysis that shows the marks from all other judges when the judge in question has a mark or score boxed by FPS as an anomaly. This adds a cross-panel insight for each judge where they have been assessed as significantly different to the other judges, and when taken in conjunction with the Form-B/C figure can help the judge to interpret why their view has been discarded.

Flying Orders

This is a key area in the operational use of the software, and requires a rapid balance between results calculations from the immediately preceding sequences and accommodating multiple aeroplane use by various pilots. Depending on the contest situation, the Jury may also intervene to vary or cut the normal lower to higher-scoring pilot grouping system to meet local needs. Further work on ACRO is in hand to provide a more flexible capability to meet these high-pressure situations and allow the scorer to move quickly toward collation of the judges paperwork for the ensuing sequence.

Warm-up pilots will also be added to the ACRO flying order above the contest pilots and without affecting their numeric references, an omission that caused some awkwardness at WAC.

The Role of our "Contest Scoring Software"

In the past the contest scoring system has been a largely unseen affair relegated to backroom specialists, expected simply to provide calculated results on demand. The scoring office however is now required to provide speedy reporting streams for competitors, judges, jury members and the contest management team in text and pictorial formats via both hard copy and internet routes, and to service the increasingly important media requirements at all our events. Post-sequence analysis for pilots and judges to reflect on their performances is another key area that will continue to drive the contest software package.

I am pleased that ACRO appears to have satisfactorily met this requirement during the 2009 season, both at CIVA championships and at many other smaller national events around the world. I am concerned however that it's operational content and style remains entirely of my own making. In corporate situations managing key IT systems is a boardroom responsibility to match planned requirements and user patterns, and I would suggest that it is in CIVA's best interest to task an action group with formalising clear contest IT guidelines for the future. This would ensure that our information systems match upcoming and published rule changes, and help to identify key areas for broader development – for example online judge data handling, wider use of the internet to link contest operations and results to satisfy growing media requirements, and possibly adaptations of our statistical systems for more 'instant' scoring output.

Nick Buckenham
ACRO software author
9th September 2009