Connected Places Catapult

MUSICC Symposium 2019

Outcomes from interactive sessions

August 2019

CATAPULT
Connected Places

Executive Summary

As Highly Automated Vehicles (HAVs) become more widespread, regulators will need to ensure that they do not pose unacceptable risks to the public. One of the big challenges of HAV testing is achieving enough coverage of real-world situations to know that Automated Driving Systems (ADSs) are acceptably safe before deployment. It is now widely accepted that testing on public roads alone is not sufficient: too many miles need to be driven to achieve meaningful results, and some important variables (e.g. behaviour of members of the public) cannot be controlled. Therefore, it is likely that simulation will be used extensively for testing HAVs.

MUSICC is creating a catalogue to store and share a library of scenarios: an ADS will be expected to demonstrate its performance in these before release to market. In the first phase of the project, a functional prototype system was built. The symposium is part of the second phase of the project, which involves gathering feedback from real users and understanding how test scenarios might be used in a future regulatory regime. The aims of the symposium were to facilitate general knowledge sharing, gather industry views on regulation and testing, and gather information to inform further development of the MUSICC catalogue system. As well as presentations by invited speakers, three interactive sessions were held during the symposium. These addressed the inputs, internal functionality, and integration of the catalogue into a wider regulatory framework.

The first session (inputs) focussed on how scenarios for the database should be gathered. Participants were invited to suggest, rate the importance of, and discuss practical/commercial barriers to, different methods of scenario collection. The results show that a combination of methods will need to be used: relying on a single source (e.g. submissions from Automated Driving System developers) is unlikely to result in a sufficiently comprehensive catalogue. Commercial incentives will need to be designed to encourage a range of organisations to share their expertise in this area.

Following a demonstration of the prototype system, the second session addressed future requirements for the catalogue system's functionality. Attendees were asked to rank features in priority order, and outputs from this session will inform the next phase of development.

The third and final interactive session addressed the wider regulatory framework for HAV approval. Four topic areas were selected for this session, these were:

- Ensuring confidence in simulator tools. The results from this part of the session showed that a method of checking simulator accuracy needs to be created, and that a formal accreditation process for tools may also be required.
- The balance of simulator and real-world testing. As expected, the group showed strong support for doing most testing in simulation. However, the results also indicated that there is not yet a consensus on what other testing should be done.
- How pass/fail criteria should be specified. It was agreed that pass/fail criteria could be defined using a performance envelope specified with each scenario, but the process to decide what this envelope should be was not clear.
- Whether perception (identifying speeds and positions of objects) should be tested separately to decision making. No consensus was reached on this topic, which has important implications for populating the database with scenarios.

Contents

Notice

CPC assumes no responsibility to any other party in respect of or arising out of or in connection with this document and/or its contents.

This document has 32 pages including the cover.

AUTHORISATION:

ACTION	SIGNATURE BLOCK	NAME AND POSITION WITHIN CPC
Written by:	Rob Myers	Technologist
Reviewed by:	Zeyn Saigol	Principal Technologist
Authorised by:	Richard Holland	Principal Technologist

RECORD OF CHANGES:

REALEASED TO	VERSION	REASON FOR CHANGE	DATE
MUSICC Website	1.0	Initial release	08/08/2019

1. Introduction

Project overview

As highly automated vehicles (HAVs) are deployed on public roads, regulators need to make sure that these systems are safe. This challenge requires coordination between regulators and system developers – with common (or at least aligned) approaches to validation desirable.

One of the big challenges is achieving enough coverage of real-world situations to know that Automated Driving Systems (ADS) are acceptably safe. Automated Driving System (ADS) developers have indicated that testing by driving vehicles on public roads is not enough, due to:

- The large number of miles needed to cover all situations which could be encountered.
- A lack of control of the test parameters.

The solution is likely to involve more than one technology but will almost certainly include an element of simulation.

MUSICC is creating a system to store and share scenarios: an ADS will be expected to demonstrate its performance in these before release to market. As far as we know, MUSICC is the first project of its type to build a proof of concept system specifically designed for regulatory use.

The project is led by a regulator (Department for Transport) and an impartial, neutral mediator (Connected Places Catapult). It is influenced by a highly credible industrial advisory group, including representatives of several major OEMs, tool providers, ADS developers, research and innovation organisations, CAV testbeds and insurers.

A working prototype system has been built and is being trialled by interested stakeholders. The symposium forms part of the second phase of the project, which involves gathering feedback from real users and understanding how test scenarios might be used in a future regulatory regime.

Symposium scope and objectives

The MUSICC Symposium brought together representatives from government, industry and academia to discuss how type approval for HAVs could work. The aims were to facilitate knowledge sharing, gather industry views on regulation and testing, and gather information to inform further development of the MUSICC database. These were addressed through three interactive sessions on the following topics:

- Inputs - how scenarios to be included in the database should be identified and selected.
- Process - the feature requirements for the database itself.
- Outputs - the wider regulatory context around MUSICC.

To help keep discussions focused, the following topics were defined as out of scope for this symposium:

- Comfort and performance issues
- Level 3 and handover issues
- ODD bounding issues
- Ethics
- Terminology and definitions

- Connectivity (V2V, V2I, etc)
- In service safety management (e.g. technical inspections, operational safety cases)

Methodology and Event Structure

The symposium included a mixture of presentations by industry speakers and interactive sessions to gather feedback from attendees. Presentations were selected to contribute to general sharing of knowledge and to provide context for the interactive sessions. Interactive sessions were held in smaller groups, with approximately 5 attendees and a moderator in each. Table 1 below shows the structure of the event.

First, three presentations were given which aimed to set the context for MUSICC, explain its role and the regulatory context. This was followed by an interactive session on the methods of, and barriers to, acquiring and sharing scenarios for regulatory use. After a break, the group returned to a presentation about assessing ADSs against a Digital Highway Code. This was followed by a demonstration of the prototype system and possible options for future development. A second interactive session asked the groups to rank these options by priority.

The third set of presentations explored practical methods of testing vehicles. Finally, the last interactive session asked groups to discuss one of the questions below:

- How do you ensure fairness of / confidence in simulator tools?
- How do you apply objective pass/fail criteria?
- What mixture of simulator/ closed road/ public testing is needed?
- Does perception need to be tested separately? What does this require from the ADS?

Interactive sessions were facilitated by a moderator on each table and included an element of online voting to gather quantitative data[1]. Detailed descriptions of the interactive sessions are included in the relevant chapters.

[1] Using the system provided by https://www.mentimeter.com/

Table 1: Structure of symposium event

Structure of Symposium event
Donald McDonald, representing the UK Department for Transport, set the context for the project by outlining the UK's regulatory agenda for HAVs
The challenges of regulating automated vehicles Connor Champ gave an update on the Law Commission's work regarding the regulatory framework for connected and autonomous vehicles.
Vision for MUSICC Zeyn Saigol, technical lead for MUSICC at CPC, presented the project scope, context and objectives
Interactive session: Acquiring and sharing scenarios
Towards a Digital Highway Code Iain Whiteside presented FiveAI's work on defining a Digital Highway Code to test HAV performance against.
Matthew Coyle, part of the CPC software engineering team, demonstrated the functionality of the MUSICC system
Interactive session: What could MUSICC do for you?
The role of physical testing in the CAV engineering lifecycle Tim Edwards, of Horiba Mira, discussed the various techniques and facilities available for Connected and Autonomous Vehicle (CAV) testing
AV systems validation using SCANeR™studio simulation SW Alex Grandjean and Bocar Sall gave a presentation on AV Simulation's capabilities for virtual testing of highly automated vehicles
Interactive session: How could type approval for HAVs work?

Attending organisations

No restrictions were placed on the individuals or organisations attending the symposium. 36 organisations were represented, including OEMs, simulation tool providers, ADS developers and independent research organisations.

Key definitions

The terms and definitions below were used throughout the symposium and in this report. This is not intended to be a comprehensive list or to create standard definitions: other projects are working on this.

Term	Definition
CAV/HAV	Connected or Autonomous Vehicle/Highly automated vehicle
Level 1-5 (SAE L1-5)	Levels of autonomy as defined by SAE J3016 (https://saemobilus.sae.org/content/J3016_201806/)
Scenario	A test case for an automated driving system
Functional scenario	A description (possibly, but not necessarily, in natural language) of a scenario's features
Concrete scenario	A machine-readable scenario where all parameters are fully defined
Logical scenario	A machine-readable scenario where some parameters are defined only by a probability distribution
ADS	Automated Driving System. The software controlling a L3-L5 CAV/HAV
Ego vehicle	The vehicle which is under test in any given scenario
Metadata	Data which describes the content of a scenario. Can be used to search for relevant scenarios in the MUSICC database.
Perception	The process of translating the real world, as observed by sensors, to a model which is useful for decision making (likely including environment features, obstacles, and road user movements)
OpenDRIVE	An open standard for a data format used to describe road network features
OpenSCENARIO	An open standard for a data format used to describe road user movements in a scenario
Operational design domain (ODD)	Defined criteria for where and when an ADS should function. This may include restrictions on weather, road types, geo-fencing, and other aspects
Scenario Description Language (SDL)	A machine readable, language to define the content of a scenario. Includes description of static content, dynamic content and metadata. For the MUSICC database, this takes the form of OpenDRIVE, OpenSCENARIO and MUSICC XML.
Type approval (TA)	The confirmation that production samples of a design will meet specified performance standards

2. Session 1: Scenario identification

Overview of issues

An appropriate set of scenarios is essential for any scenario database to be useful. For regulatory use, these need to capture the reasonably foreseeable risks associated with HAV functionality, without presenting an unreasonable barrier to the market. This means that the scenarios need to be comprehensive, high quality, and represent a fair test.

There is not yet a consensus about how such a database should be populated. Understanding the methods likely to be used is important for determining the functionality required from the database and the actions required by the regulator. This session gathered the views of attendees on scenario identification.

Format of session

Four types of technical approach to gathering scenarios were identified before the symposium, as summarised below:

- Real driving by test vehicles: any approach involving gathering scenarios based on data from vehicle-mounted sensors on the public road. This includes data from human driven vehicles (which may or may not have an ADS working in shadow mode), pre-production HAVs and (in future) deployed HAVs.
- Crash/ telematics data: approaches which involve gathering incident or near-miss data from human drivers in ordinary cars. This includes crash databases such as STATS19 and GIDAS.
- Theoretical/ top down risk assessments: possible crash scenarios identified and evaluated using a structured risk assessment methodology.
- ADS developer knowledge and experience: known challenging scenarios submitted or suggested by ADS developers.

Participants were asked to discuss, then score using an online voting tool, the relative importance of these approaches to scenario collection. They were also invited to suggest any which were missing from the list. A follow up exercise explored the practical and commercial barriers to each approach in more detail.

Results – relative importance

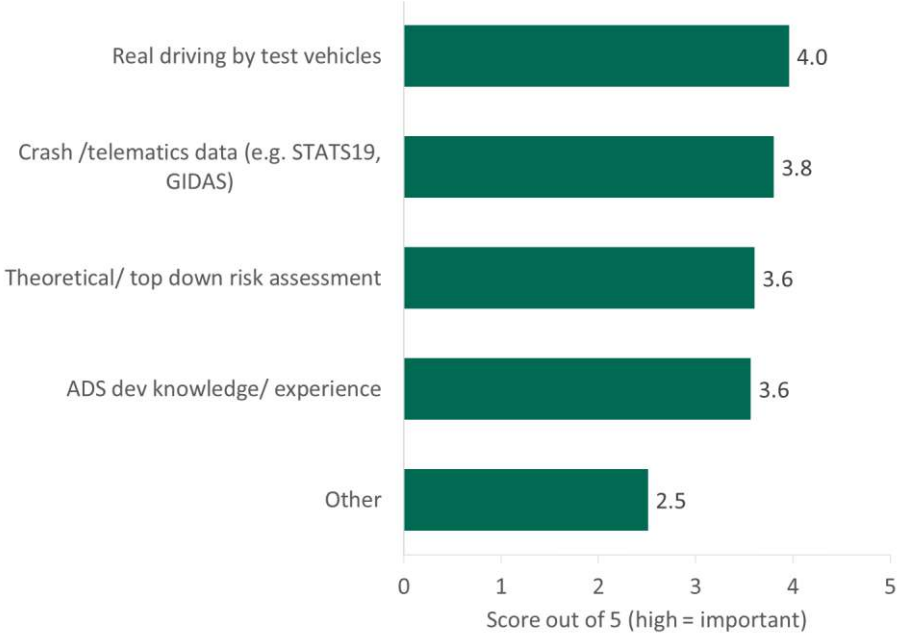

Figure 1: Voting results for relative importance of scenario collection approaches

Figure 1 above shows the voting results for different ways to gather scenarios. All four of the suggested approaches received similar ratings, with 'other' receiving a much lower rating. When combined with the qualitative comments, this suggests that:

- The four broad categories of method previously identified are all quite important.
- For a regulatory database, relying on one alone is unlikely to be enough.
- These categories cover most of the important techniques currently known to the industry.

Results – other possible approaches

Several attendees suggested additional methods of gathering scenarios, as summarised below:

- Gathering data directly from an ADS. An automatic process would identify scenarios where the ADS performed poorly and mark these for future use.
- Crowdsourcing suggestions from members of the public, perhaps using a gamified process.
- A large-scale simulation to test the interaction of ADS and human drivers.
- Similarly to the above, carrying out physical tests on a proving ground to observe user interactions with the ADS.
- Real world data can be gathered from static sensors, as well as test vehicles.

Discussion of the effectiveness of these occurred during the 'barriers and solutions' sub-session.

Results – barriers and solutions

This section explores the barriers to collecting scenarios using the methods outlined above and their implications for a regulatory scenario catalogue. Any solutions proposed by the symposium participants to overcome these barriers are also included.

All methods

Table 2: Barriers to collecting scenarios which apply to all approaches

Barriers	Solutions proposed by participants
International compatibility: behaviours and roads vary between locations, as do environmental factors	Include location/driving culture specific scenarios
	2-stage testing approach: fundamental tests valid worldwide, supplemented by location specific tests which consider environment and culture
Need for common structure/language/standards	Scenario database should be structured in a way which enables attributes/characteristics to be added as the need for new attributes is identified
	Need for international standardisation efforts
Conversion from functional or concrete to logical level may be technically challenging	
Funding for scenario development (especially applies to accident data and top down risk assessment approaches, where there is no obvious commercial incentive to develop and share scenarios)	
Whoever creates scenarios will (unintentionally or otherwise) bias what is included based on their previous experience.	Several different approaches to scenario collection should be applied. Involve people from a variety of backgrounds in the process.

Implications for regulatory database
• Localisation needs to be considered as part of the operational design domain. It should be easy to identify which scenarios are appropriate to a location using the metadata. It may be useful to identify localised variants of the same scenario. • Consideration needs to be given to the process for adding new attributes to metadata, as our understanding of requirements changes • International standardisation is crucial, and the MUSICC team should remain engaged with standardisation efforts • The database will need to store scenarios generated through several different methods

Real driving by test vehicles

Table 3: Barriers to collecting scenarios through real driving by test vehicles

Barriers	Solutions proposed by participants
Critical events occur rarely in normal driving	Use machine learning approaches to extract the scenarios which are likely to be a challenge for an ADS
Scenarios seen as intellectual property by developers	Require data to be anonymously shared (e.g. US approach).
	Only share a general description of the scenario, not a full, ready to run, specification
Legal penalties as a result of sharing near miss/crash data	Free agreement to anonymously share near miss data.
	Provide appropriate incentives.
Inconsistent data format of outputs from vehicles	A uniform format could be required (analogous to OBD2 or OSI standard)
Incorrect sensor outputs could be included in scenario	Include some means of determining ground truth (e.g. superior sensor set for verification, camera)
	Test in 'shadow mode', where the ADS is running but does not control vehicle
Sharing scenarios could expose commercially sensitive information about the ADS which collected them	Investigate the mechanisms by which this could happen and the level of risk which they present.
	Only share a general description of the scenario, not a full, ready to run, specification
Can be expensive, both to collect and to compute	
Lots of driving required to achieve enough coverage (weather, culture, signage). May be difficult to know whether this is complete.	
Testing likely to use professional test drivers, so not representative of ADS behaviour	

Implications for regulatory database
• Real world test drives appear unlikely to be sufficient as the sole source of scenarios • Need to relate logical scenarios in the database to general descriptions (i.e. functional scenarios) generated elsewhere

Crash/telematics data

Table 4: Barriers to collecting scenarios using crash data or data from vehicle telematics systems

Barriers	Solutions proposed by participants
Crash data has a lack of structure or insufficient detail	Regulators are often the experts on crash data and may need to generate scenarios from it themselves
	Specialist accident investigation board (similar to UK AAIB/RAIB/MAIB) to collect data from AV crashes.
	STATS19 should be updated and augmented to be more useful for scenario generation (3D area scan, touchscreen instant re-creation)
Labour intensive to interpret	Make automated interpretation possible – see above suggested improvements to crash data
Different collection methods/data formats in each country	Regulators are often the experts on crash data and may need to generate scenarios from it themselves
Privacy (both crash data and telematics)	Careful anonymisation process required
Existing crash data reflects human failures, AV failures may be different	Combine with other collection approaches
	Look carefully at any future AV crashes
Information on interaction between human drivers and AVs is not available yet	Real time simulation with both AVs and human drivers
Data accuracy - not fully objective	
Telematics data is not detailed enough, not shared	Create a market for telemetry data, so there is a commercial incentive to collect required data
Telematics IP issues	Regulation for data sharing
Telematics lack of standardisation	Create standards

Although it was not specifically requested at the start of the session, participants also noted the following barriers and solutions relating to crash data from autonomous vehicles:

Table 5: Barriers to collecting scenarios based on real life AV crashes

Barriers	Solutions proposed by participants
Lack of structure or insufficient detail	Mandatory data storage system for AVs
ADS developer action – deleting or refusing to share data which may show liability	Impose legal requirements for maintaining data integrity

Implications for regulatory database
• Integration with crash databases may be desirable. Users may wish to create scenarios based on crash data or to check whether existing crash scenarios are represented. • When an AV crash occurs, it should be possible to identify relevant scenarios from the database (to understand whether there is a gap in the scenario set which needs to be addressed).

Top down/theoretical risk assessment

Table 6: Barriers to collecting scenarios based on a risk assessment approach

Barriers	Solutions proposed by participants
Scenarios limited by assessors' imagination	Working group for assurance with assessors from varied backgrounds
	Fit into safety case approach
	Use in combination with other methods of scenario collection
	Crowdsource scenarios (gamification to allow members of the public to submit them?)
Tendency to identify human collision causes	
Commercial/IP: risks identified by ADS developers could give insights into system design	
Agreeing on behaviour models of other actors	Risk assessment should attempt to identify everything which people are likely to do

Implications for regulatory database
• Requirement to be able to use risk assessment outputs to identify missing scenarios

ADS developer experience

Table 7: barriers to using ADS developer expertise to identify scenarios

Barriers	Solutions proposed by participants
IP – lack of commercial incentive to share	UN ECE already provides a forum for information sharing
	Regulator needs to have (and pay for) its own expertise. Could come from own staff, research organisations, academia, and technical consultancies.
	Regulator needs to force developers to share scenarios as part of type approval process
	Only share a general description of the scenario, not a full, ready to run, specification
	Explain potential commercial advantage of sharing scenarios which their ADS can pass – allowing competitors to deploy unsafe vehicles could damage the market for HAVs
Incentive to delay more advanced competitors through adding unmanageably difficult scenarios	Quality filter for regulatory adoption. UN ECE/DfT/3rd party approval required.
ADS developers are likely to miss 'unknown unknowns' – if they had known a scenario existed, the ADS would already be designed to cope with it	Supplement with other sources
ADS developers may only contribute scenarios which are known to pass **Example from telecoms industry: a test is only acceptable once several OEM's products are known to pass it (limits scope for using OEM scenarios)**	Combine with other approaches

Implications for regulatory database
• Needs to support a scenario approval workflow to ensure quality

Discussion of other suggestions

Table 8: Discussion of other suggestions

Suggestion	Discussion
Gathering data directly from an ADS	This has the potential to allow lots of scenarios to be gathered quickly. However, this approach has all of the barriers of using ADS developer experience, plus extra technical challenges presented by the large volume of data.
Crowdsourcing suggestions from members of the public	Many members of the public are very interested in automated vehicle development and may be able to suggest scenarios which others may not have thought of. This approach would require a system which is fun to use, has widespread uptake, and produces useful outputs.
Large scale simulation, testing interaction of ADSs/ human drivers	This is the only approach suggested which allows real interactions between human drivers and ADSs to be tested. It suffers from some of the same problems as real world tests: a lot of driving is likely to be required to find challenging scenarios.
Physical tests on a proving ground, testing interaction of ADSs/ human road users	This approach makes fewer requirements on simulation infrastructure than the one above, but otherwise has similar characteristics. Large-scale controlled environment testing is likely to be expensive.
Gathering real world data from static sensors	When compared to using test vehicles, this approach is likely to result in more observations but from a limited geographic area. As there is no Ego vehicle to take direct measurements from, the data produced may be different.

Implications for regulatory database
All the suggestions above could be useful for gathering scenarios, though the low score given to 'other' in the voting suggests they may not be essential. These approaches do not create significant additional requirements for a scenario database.

3. Session 2: What could MUSICC do for you?

Overview of issues

CPC is planning to further develop the MUSICC database system's functionality. This session gathered views of attendees on what the priorities for feature development should be.

Format of session

This session followed a demonstration of MUSICC's existing functionality. Users were asked to discuss, then rank in priority order, the following development options:

- Enhancing metadata to more precisely specify a vehicle's Operational Design Domain (ODD).
- Providing a user installable package, allowing organisations to create private versions of the database. This would include functionality to easily fetch scenarios from the master database while allowing private scenarios to be stored in a MUSICC-compatible format.
- Representing functional scenarios (abstract, human readable, descriptions of scenario content) explicitly in the database.
- Integration with other tools used in the scenario creation/ simulation workflow.
- Automated generation of images to represent scenario content. At present, any images to be shown must be uploaded with the scenario.
- Some form of user rating, to highlight scenarios which are especially valuable or challenging.

Moderators at each table recorded the reasoning behind the rankings and any other suggestions made by their group.

Results

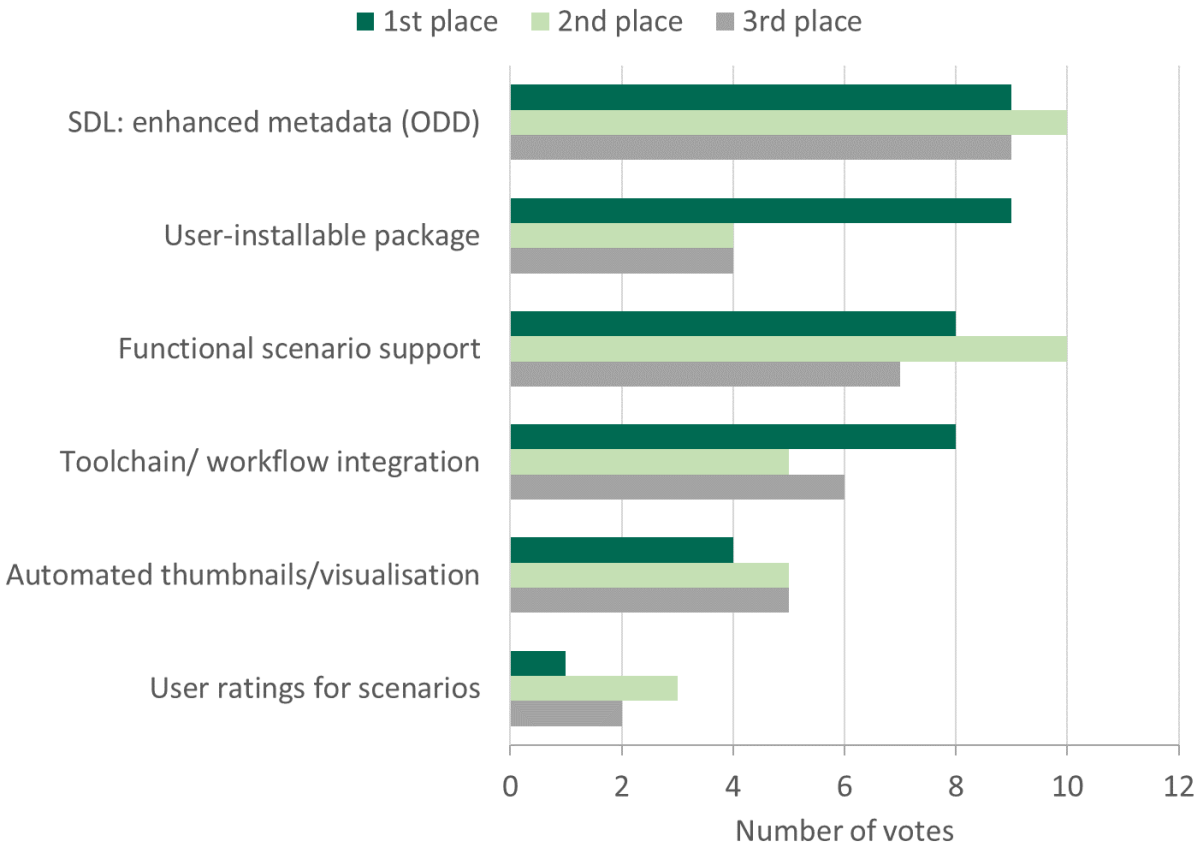

Figure 2: Development options prioritised by attendees

Figure 2 above shows the number of voters who ranked each option in first, second or third place. Several of the development options recieved similar results.

Enhanced metadata to describe the ODD was the overall favourite, receiving 9 first preference and 10 second preference votes. However, the "user installable package" and "support for functional scenarios" options were also popular. "Toolchain and workflow integration" divided opinion: while 13 delegates listed this as one of their top two areas for development, 10 listed it in their bottom two. "Automated visualisation" recieved intermediate scores, while "user ratings for scenarios" was clearly not seen as a priority by the group as a whole, with only one attendee putting it in first place.

The following subsections summarise the reasons given for ranking options in this order. They also document additional suggestions from the group.

SDL: enhanced metadata

Enhancing the metadata was a priority for the group, with a large proportion of users listing it as one of the most important features. The following comments were noted by moderators:

- It is important to be able to search separately for static content (e.g. road network features) and dynamic content (e.g. vehicle manoeuvres). These should be separated in the metadata.
- Metadata should include more attributes, such as reflectivity of surfaces. This could, for example, be relevant in wet conditions. However, it is necessary to consider which elements should be specified in the meta-data and which should be included as dependent variables of another high-level parameter (e.g. weather conditions).
- While the current metadata includes authorship information, it now seems likely that scenarios will come from more than one type of source (e.g. all of the approaches discussed in Section 2). Metatdata should include more detail about the scenario's origin.
- Metadata needs to be designed in a way which is flexible, to accommodate future changing requirements.

User installable package

The idea of providing a user installable database system divided opinion. It received a high proportion of first place votes, but very few attendees ranked it in second or third place. The following comments were noted:

- An installable package available would help to drive adoption of the standards behind MUSICC, making the regulatory process simpler.
- Creating private versions of the software should not be the role of the regulator, this should be left to the market. Releasing private versions could detract attention from main regulatory database.
- Some participants were indifferent to this suggestion, since their organisation has no requirement to store scenarios.

Functional scenario support

Support for functional scenarios was a relatively popular option, with a high proportion of users listing it within their top three choices. Relatively few comments on this topic were noted by the moderators, suggesting that the uses are not yet clear. Some attendees suggested that the regulator should specify functional scenarios first, then ensure that the test set contains these as a second step.

Opinions on toolchain and workflow integration

Users expressed an interest in integrating MUSICC with other systems for scenario creation and simulation. However, this was considered less important than the features above. Comments include:

- Integration with a simulator tool would be valuable to drive adoption and demonstrate the system.
- Integration with simulator tool could be seen as endorsing a particular simulator, and should be left to the developers.
- Proving that the API is useful for another tool is important and should be demonstrated.

- Tools to support scenario creation would be valuable. This could include integration with local authority datasets, crash data, or other databases (e.g. TNO Street Wise).
- A structured scenario upload process is required.

Automated visualisation

Attendees ranked this feature as a low priority. It was suggested that automatically generated images would be more consistent, and this might help users to navigate the database. However, most scenario authors will view their scenarios before upload, so producing images manually is straightforward.

User ratings for scenarios

The fewest high preference votes were received for applying user rankings to scenarios. Moderators noted the following comments:

- User ratings would not help regulators to determine which scenarios are important
- Ratings may have value in understanding whether scenarios contain appropriate behavoiurs for local cultures / driving styles.
- User needs are diverse, so different users will have different opinions on which scenarios are valuable. This means the ratings are unlikely to be meaningful.
- May have applicability to a crowdsourcing approach (see Section 2).

Other feature suggestions

The following other features were suggested:

- Support for OpenSCENARIO 2.0 when it is released
- Inclusion of pass/fail criteria
- Features to support checking for duplicates
- Provision to upload anonymised result data (e.g. ADS pass/fail). This could be useful for understanding which scenarios are useful and which are poor quality, but developers might be reluctant to provide these results.

4. Session 3: How should type approval for HAVs work?

Overview of issues

Development of the MUSICC database is intended to generate knowledge about how a future regulatory process for Highly Automated Vehicles could work. Building a proof of concept system is part of an approach of 'learning by doing': demonstrating practically some of the tasks which must be done as part of type approval. We intend this understanding to be used to inform future initiatives in the area of CAV certification.

The final session explored some of the topics around how MUSICC might be used, and what other components will be required by a type approval process. It aimed to identify areas where there is an industry consensus for a particular approach, and areas where one still needs to be built. The outputs are findings in their own right, but also help to provide context system development to be undertaken during Phase 2 of the project.

Before the workshop, four key questions to be addressed were identified. These are listed (with some contextual description) below:

- How do we ensure fairness of/ confidence in simulator tools?

  It is anticipated that a simulator test will need to be passed as part of a regulatory assessment. This makes it important that the software used is a reasonable reflection of reality, and that developers cannot 'game' the test by using their choice of software.

- What mixture of simulator/ closed road/ public testing is needed?

  Almost all organisations working on ADSs believe that simulation will be an important part of the testing process. This question aims to understand the other testing that will be required, and what simulation cannot achieve.

- How do you apply objective pass/ fail criteria?

  There needs to be some mechanism for deciding whether performance on a test is good enough. This may require performance expectations to be specified in the scenario database.

- Does perception need to be tested separately?

  Some ADS architectures draw a distinction between perception (identifying and analysing objects in the world which are relevant to the ADS) and decision making (planning a route based on this information). The decision on whether to test the performance of these aspects together or separately has substantial implications for both a scenario database and the regulatory system around it. If these systems are tested together, as a single device, the scenarios become more complex. It is necessary to include enough information for the scenario to be reconstructed by virtual sensors, which may require highly realistic textures and 3D models. In contrast, if perception is tested separately, ADS developers need to allow an appropriate interface in their software.

Format of session

Attendees were split into small groups, with at least two groups addressing each question. They were then asked to suggest answers to share with all attendees. A voting exercise was held, again asking users to rate the importance of each answer. This was intended to provide an indication of which suggestions represent a consensus position.

Ensuring confidence in simulator tools

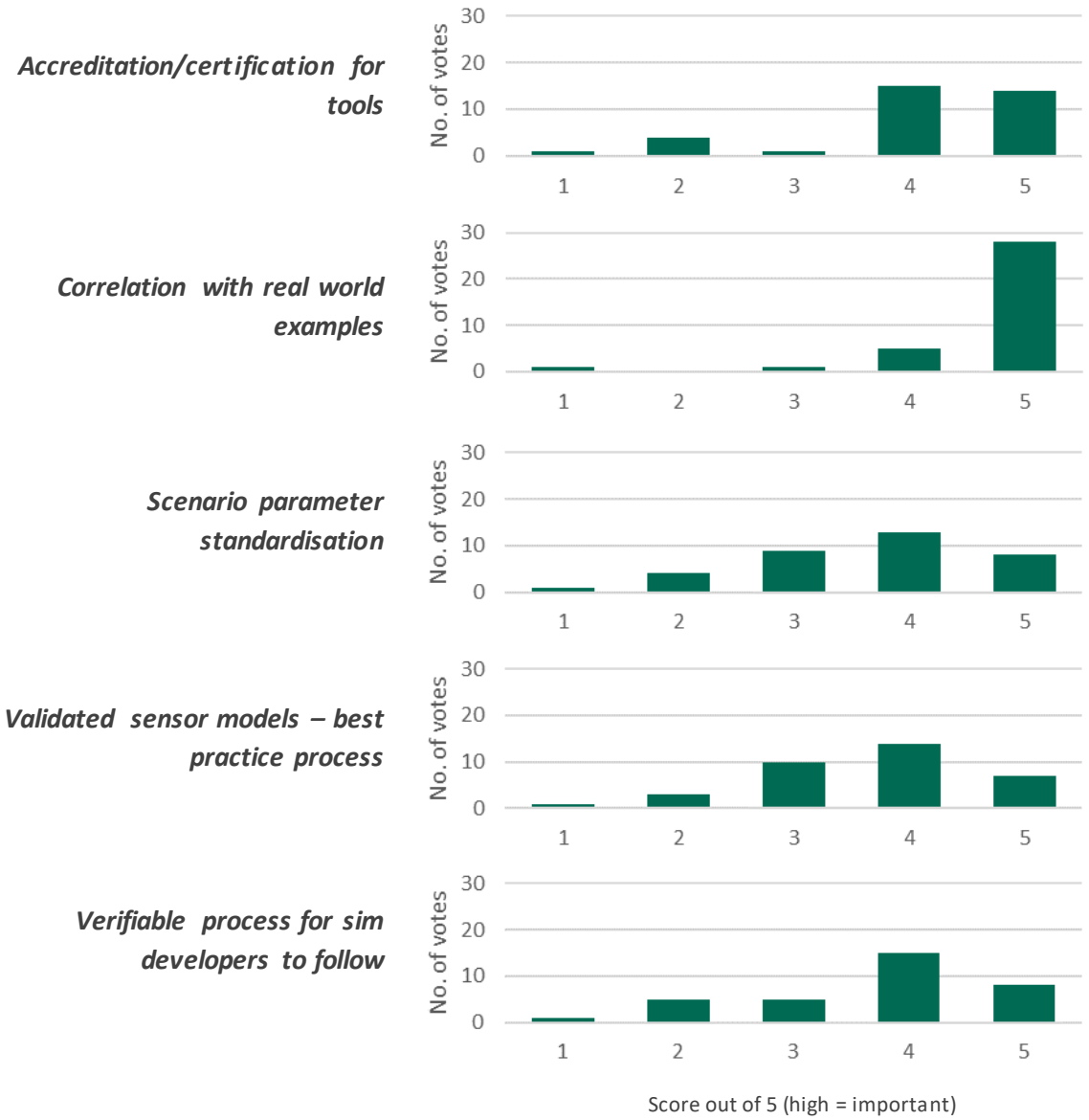

Figure 3: Voting results for methods of ensuring confidence in simulator tools

Figure 3 above summarises the group's views verifying simulator tool functionality. Five suggestions were made, as listed below:

- One group suggested that there should be a process of accreditation for simulator tools. Any tools used in a regulatory context would have to be approved by a relevant body. This suggestion was reasonably well supported by other members of the group, with most rating it at 4 or 5. However, a notable minority gave it a score of 1 or 2: this suggests that any proposal to accredit simulators may meet some opposition. The idea of defining a single regulatory simulator was

suggested and rejected, as it was thought to constrain ADS developers too much. Specifying a simulator would mean that only features supported by that sim tool could form part of the ADS (e.g. a novel sensor type could not be used). Also, it may be technically difficult or impossible to integrate the ADS developer's system with a regulator-specified tool.

- The least controversial suggestion was to check that simulation results are consistent with results in the real world. All but two individuals gave this a score of 4 or 5. It is worth noting that simulation will never be perfect: a decision will need to be made about how much deviation from the real world is acceptable. A new standard could be required to specify this.
- It was suggested that some parameters or models used by the tools may need to be standardised (e.g. every simulator should use the same friction coefficient). Other attendees suggested that alternative approaches (e.g. a requirement to declare this type of information) might achieve the same outcome.
- Simulator developers could be required to follow a structured, verifiable tool development process. This would help to produce consistent quality tools. In the discussion session, some participants suggested that it also important for ADS testers using the tool to follow a defined process. Simulator tools are often designed to be highly configurable, which means that the user can have a large effect on the accuracy of the results.
- If simulation is used to test perception, the ADS will need simulated sensor data as an input. This will require a model which predicts what the outputs from a vehicle's sensors would be. Since the accuracy of this model will affect the test results, the process of validating it is important. Best-practice guidance for this is under development.

What mixture of simulator/closed road/public testing is needed?

Figure 4 below shows the voting responses relating the balance of closed road and public testing.

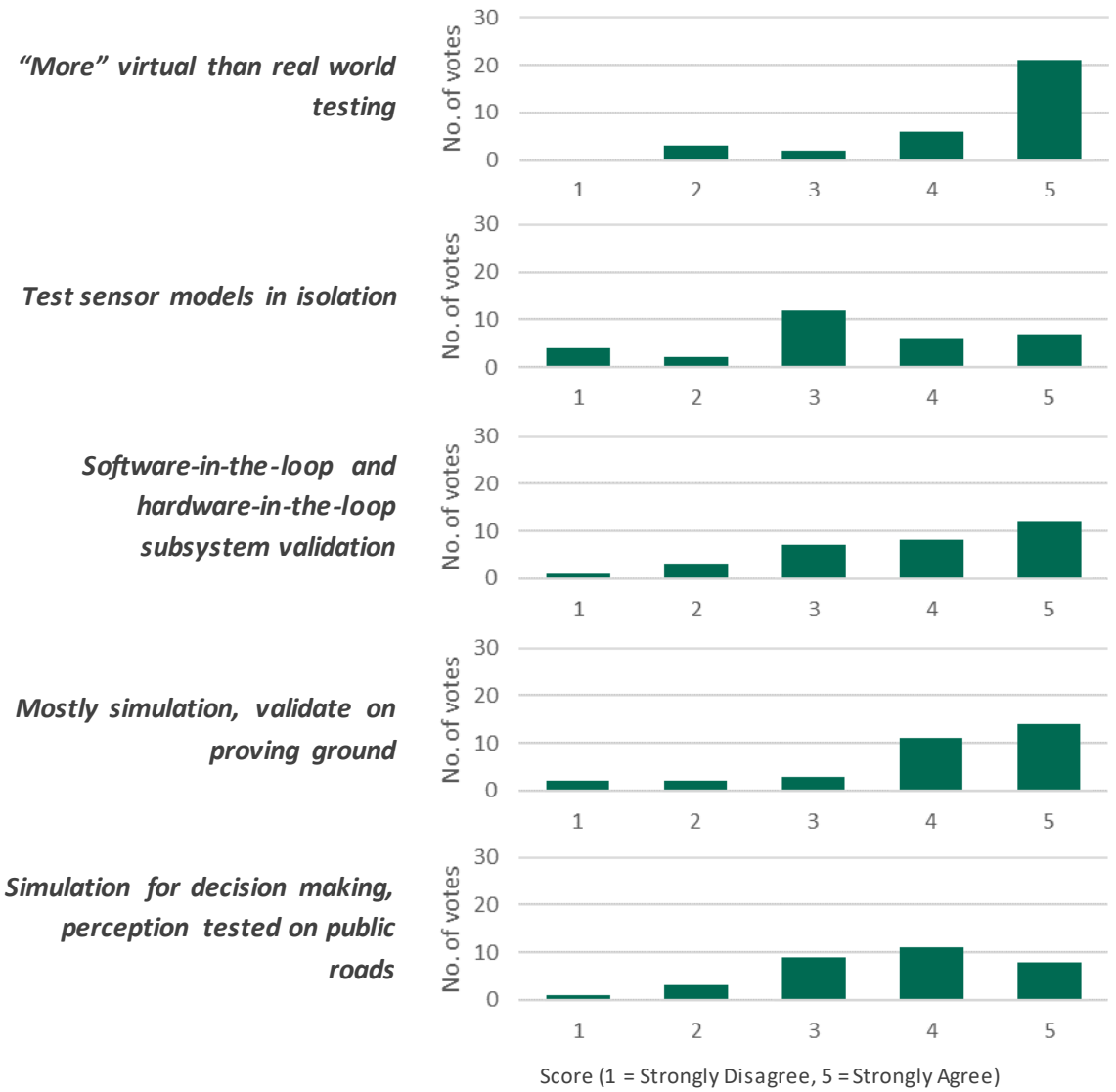

Figure 4: Voting responses about the required mixture of simulator and real-world testing

Five options were identified, as summarised in the bullet points below:

- There was a clear consensus amongst the group that most validation for ADS systems should be done in simulation. Almost all participants gave this a score of 3 or above, with more than half assigning it 5 out of 5.
- The suggestion that any sensor models used in simulation should be tested separately divided opinion. Sensor models form part of a perception system, and views differ about whether perception should be tested in simulation at all.
- It was suggested that as well as an overall ADS test, simulation could be used to test subsystems separately. These tests might include the hardware, as well as the software, to be used. In this instance, the device under test would only be part of an ADS: the operation of other parts of the system would be simulated. This idea received moderate support in the voting, though with more negative and neutral responses than some of the other ideas.

    - Using this approach for regulatory tests may be challenging. It would require the regulator to be embedded in the design process, and the tests to be varied (but kept fair) for different system architectures. Other industries, such as aviation and nuclear power, may have relevant knowledge about the most effective ways to involve regulators in the design process.

- The final two approaches were suggested by the same group, and represent alternative strategies for testing:
    - The first involves doing most testing in simulation, including perception testing. Proving ground tests would be used to validate the results. This was a relatively (but not uniformly) popular suggestion.
        - Carrying out perception tests in simulation means that scenarios must be much more detailed.
    - The second involves using simulation to test decision making capabilities only. Public road testing would then be used to test the perception system: it was suggested that this would provide a more realistic test for the perception system's ability to interpret the world. This suggestion was less popular than the first, though a significant number of attendees preferred it.
        - Perception testing requires the ADS developer to open up the outputs of its perception system
        - Decision-making tests require the ADS developer to allow the simulator to provide direct inputs to the decision-making system.
        - Splitting the two types of test means that interactions between the perception and decision making system cannot be tested.
    - Of the 32 users who responded to this question, 14 gave a higher score to the first option than the second, 7 gave a lower score and 11 assigned the same score to both. This shows that there is not yet an industry consensus on the subject.

How do you apply objective pass/fail criteria?

Figure 5 shows the four approaches to setting pass/fail criteria which were identified by the groups:

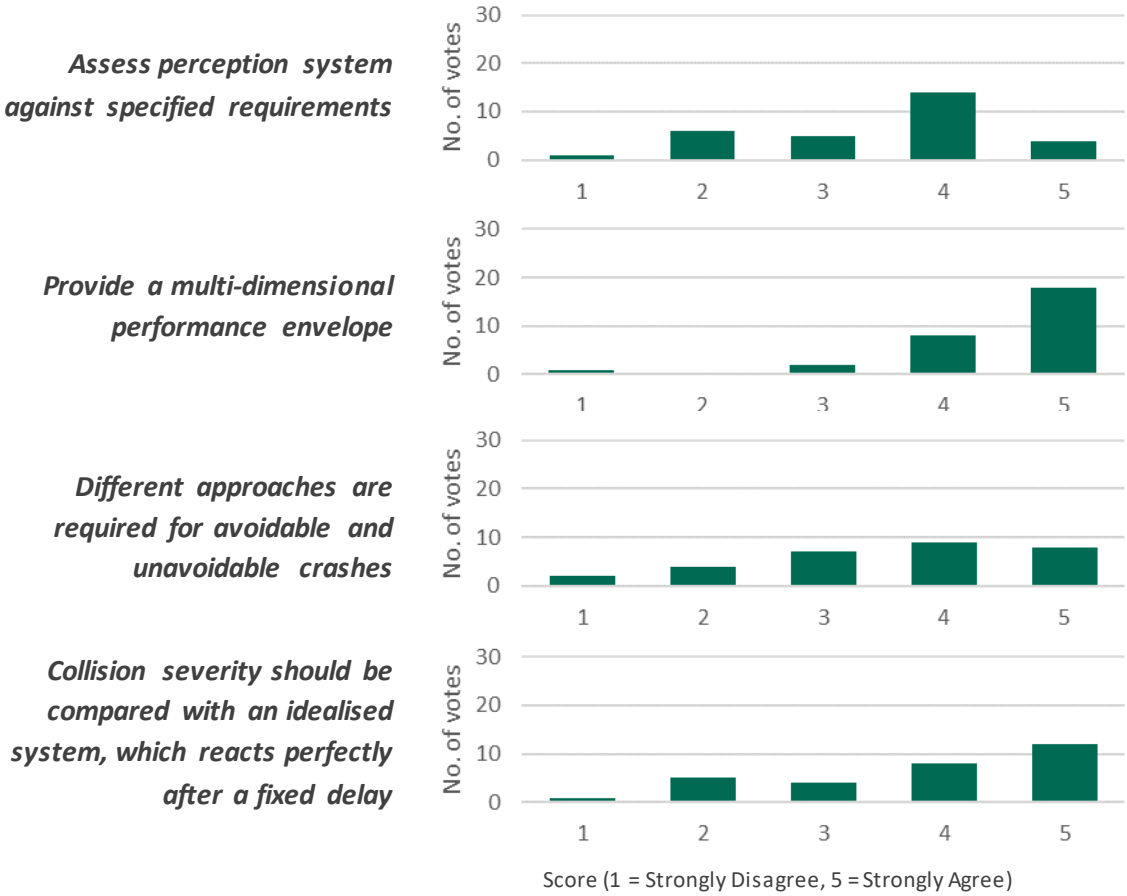

Figure 5: Voting responses for pass/fail criteria

These approaches can be summarised as follows:

- Perception could be tested comparing the output of the perception system to a performance specification. The group held mixed opinions about this suggestion.
- A multi-dimensional performance envelope for the Ego vehicle could be specified for every scenario. For example, when overtaking a stationary vehicle, specify a minimum and maximum distance between vehicles. This was a popular option, though it may be difficult to work out what the envelope should be and how it should be defined. MUSICC supports scenarios where parameters are randomised: different values may alter the performance envelope.
- One group suggested that pass/fail criteria should simply be based on whether the vehicle crashes. If a crash is avoidable in the scenario, it must always be avoided. If not, its severity should be minimised. However, this does not address scenarios where the vehicle takes an unsafe action but no crash results.
- To determine acceptable severity when a collision is inevitable, a standard based on acceptable reaction times could be used. A maximum acceptable severity would be determined for each scenario, based on the Ego vehicle taking the best possible action after a standard delay.

As well as the potential solutions outlined above, the following general points were noted:

- Pass/fail criteria could include a measure of how efficiently the vehicle drives, as well as whether the vehicle avoids a collision. Poor performance may indirectly result in an unsafe outcome (e.g. a dangerous overtake by another road user because the autonomous vehicle was stopped unnecessarily). Some attendees felt that measuring this should not be the regulators role: there is a strong enough commercial incentive to produce a vehicle which performs well.
- It is relatively straightforward to estimate the severity of a simulated collision, based on the type, mass, and relative speed of the vehicles involved. It is much harder to establish what, if any, severity is acceptable.
- If the vehicle under test behaves in a way which is confusing to other road users, this is an unsafe outcome. It is hard for a simulation tool to predict whether human drivers will find a behaviour confusing: a "Digital Highway Code" approach might be one way of checking for this.

Does perception need to be tested separately?

Figure 6 below shows the suggestions which the groups developed on the topic of perception:

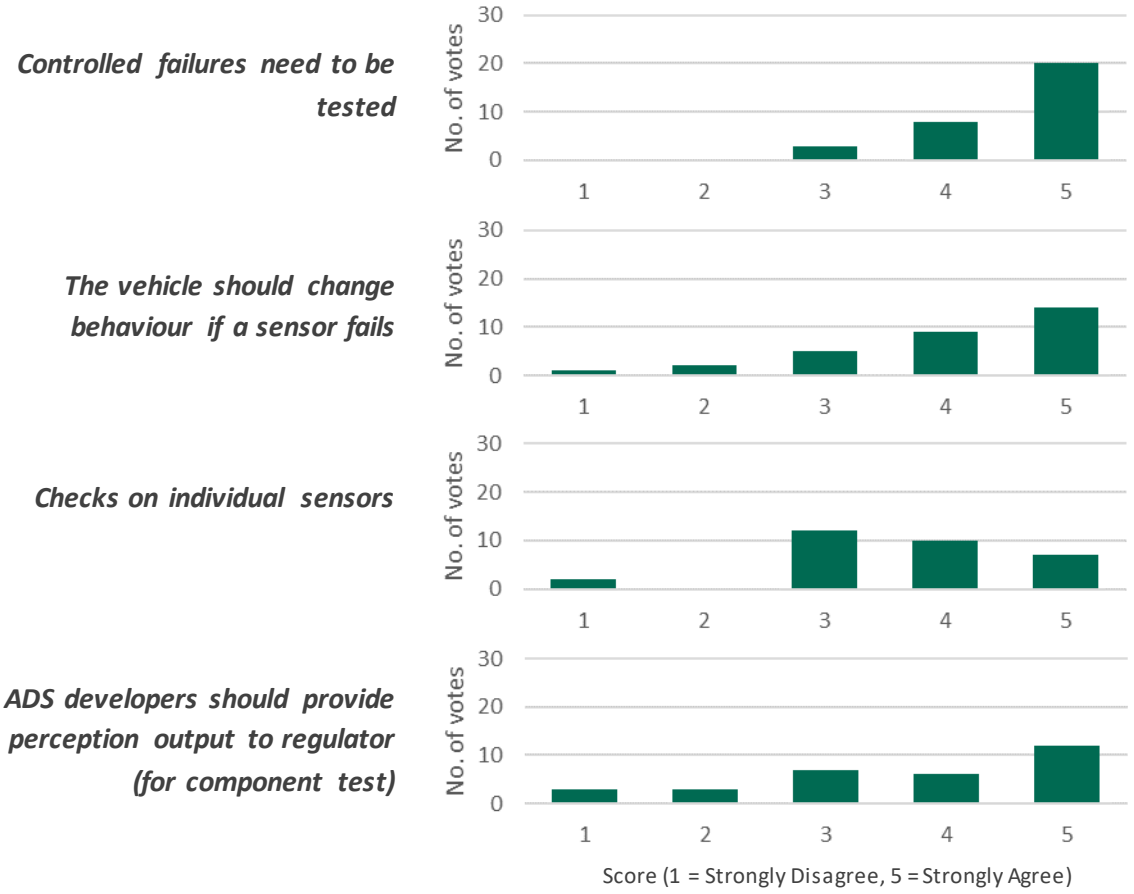

Figure 6: Voting responses for perception testing

There were some suggestions in this topic which most attendees agreed on. The first two shown in Figure 6 are an example of this: there is a consensus that sensor failures need to be tested, and that the vehicle will need to change its behaviour (e.g. execute a minimal risk manoeuvre) if a sensor fails.

Opinions were less clear about whether regulators should test individual sensors: most votes were towards the middle of the available range of values. It was noted that this approach would not allow the full perception capabilities of the vehicle to be tested.

The most controversial suggestion was that ADS developers should allow the regulator to directly observe the output of the perception system. This is probably required for the regulator to be able to evaluate the outputs of a separate perception test. This suggestion produced a wide range of scores, from very low to very high, reflecting a lack of consensus about this approach.

5. Conclusions

This section presents the key findings from the workshop and some of their implications. Outcomes which affect either the regulatory framework as a whole or the specific requirements for a scenario database have been considered.

Scenario identification

Key finding

There is no single source of scenarios which provides enough coverage for a regulatory database. Several (or all) of the following approaches will be required:

- Observation of real driver behaviour, using either vehicle or infrastructure mounted sensors
- Generating scenarios based on crashes involving human drivers
- A risk assessment carried out by appropriate experts
- Suggestions from ADS developers

While these appear to be the most important approaches now, others have been suggested which may supplement them in the future.

This finding has implications for the scenario database itself and the regulatory framework which surrounds it. Key issues include:

- The scenario catalogue system needs functionality to support the regulatory body in managing scenarios produced with different methods. Features should be considered to:
    - Track the origin of scenarios
    - Identify similar scenarios produced with different methods
    - Control the quality of scenarios entering the database.
- A variety of skills will be needed to support scenario generation. These will not all be found within one industry sector. In particular, the expertise of ADS developers is essential but cannot be relied on alone.
- All methods of scenario generation have commercial barriers which need to be overcome. ADS developers will need to share expertise and/or scenarios: for this to happen, it will have to be in their commercial interest. Similarly, other organisations (e.g. research organisations) may require a source of funding to be able to contribute.
- Further work will be required to address the technical challenges identified for scenario identification.

Database functionality

Key finding
Section 3 clearly shows the areas which attendees considered a priority for development

The session showed which features are considered important. In some cases, enough information has been obtained from the workshop to guide development. In others, further engagement with the industrial advisory group and other stakeholders will be required.

Type approval for HAVs

Ensuring confidence in simulator tools

Key findings
• Most attendees believed that simulator tools used for regulatory purposes need to be accredited or certified. • Regardless of whether certification takes place, a process is required to check that simulator results are repeatable in the real world.

Both findings above have significant implications for the regulatory process:

- If simulator tools are to be certified, a performance specification (and process for verifying that a tool meets that specification) will need to be created.
- Similarly, a specification needs to be created for correlating simulator results with the real world. This will need to consider how test cases will be selected, who will carry out the tests, and what level of performance is required to pass.

Mixture of simulator/ closed road/ public testing

Finding
• Most regulatory testing for HAVs should be carried out in simulation • There is not yet an industry consensus for how simulation should be combined with other types of test

The conclusion that most testing should be done in simulation is unsurprising but highlights the need for a catalogue of regulatory scenarios. A lack of consensus from the industry about exactly what should be tested in simulation is more problematic: understanding this is important when specifying the database content. It may be best to start by identifying requirements which are common to all, or most, uses of simulation.

Applying pass/fail criteria

Finding
• Most attendees supported the use of a multi-dimensional performance envelope to specify pass/fail criteria

There was an agreement that scenarios should specify a multi-dimensional envelope of acceptable performance. For example, a scenario which involves overtaking a parked car might specify a minimum and maximum passing distance and acceptable speeds. This creates the following requirements:

- A way of expressing pass/fail criteria in the scenario.
- A systematic way to decide what acceptable performance is for every scenario. This could be based on a digital highway code or other approach.

Should perception be tested separately?

Finding
• The idea of testing perception separately to decision making, with ADS developers providing appropriate inputs and outputs to regulators, is relatively controversial. • There may be a requirement to use scenarios to test the response of a vehicle to sensor failures.

The decision on whether regulators should test perception and decision making separately has important implications. If perception is tested separately, most scenarios will be relatively simple to define. However, if scenarios are to be used for testing both decision making and perception, highly detailed information about the world (e.g. realistic 3D models, with appropriate properties for shape and reflectivity) needs to be provided. This could mean that scenario creation becomes significantly more complex.

The MUSICC database includes functionality to support both approaches to testing. The example scenarios (provided to demonstrate the system's functionality) could be updated to include detailed environment data (e.g. 3D models of scenery) once it becomes clear what is required for testing.

CPC comment

The symposium was well attended and we highly value the efforts of all who contributed by sharing their knowledge on the day. The insights which were generated will be useful, both to the MUSICC project and the DfT, in helping to define a workable, comprehensive HAV approval regime.

Several people commented on the value of the symposium in their feedback, with some suggesting that the CPC should hold more events like this in the future. We will certainly look to do this if another opportunity arises. In the meantime, we remain open to engagement and feedback with the HAV community: please contact musicc-support@ts.catapult.org.uk if you would like to get involved.

1 Sekforde Street
Clerkenwell
London
EC1R 0BE
Tel: 020 7952 5111

The Pinnacle
170 Midsummer Boulevard
Milton Keynes
MK9 1BP
Tel: 01908 359 999

CATAPULT
Connected Places