LolaPrompts: Assisting the General Public in Performing Real-Driving Emission Tests

Melane Navaratnarajah[1], Ma'ayan Armony[1], Sebastian Biewer[2], Holger Hermanns[3], and Mohammad Reza Mousavi[1]

[1] King's College London, United Kingdom
{maayan.armony,mohammad.mousavi,melane.navaratnarajah}@kcl.ac.uk
[2] consistec GmbH, Germany sebastian.biewer@consistec.de
[3] Saarland University, Germany hermanns@cs.uni-saarland.de

Abstract. The Real-Driving Emissions (RDE) is a regulation set by the European Union Commission. Its main purpose is to set out the foundations for vehicle emission tests. In this paper, we present LolaPrompts, which exploits runtime monitoring to assist the drivers in performing driving scenarios that conform to the RDE constraints. It provides the drivers with audible prompts and explanations in order to produce a valid RDE test. The purpose of our tool, LolaPrompts, is to make RDE test available for the general public and allow them to scrutinise the emission profile of their vehicles.

Keywords: formal verification, runtime monitoring, autonomous vehicles, exhaust emission test, real driving emissions

1 Introduction

The Real-Driving Emission (RDE) test [16] is put forward by the European Union Commission to serve as a uniform benchmark for vehicle emissions across the European Union [10]. The latest specification of the test is RDE-4, which was introduced in 2019, and is the benchmark currently used in EU legislation. The regulation defines constraints on which driving style makes a test *valid* and thresholds for emissions to *pass* a valid test.

Nitrogen oxide (NOx) emission levels during vehicle approval tests can be significantly different from those of tests carried out during 'on-road driving' [11]; this could be seen through Volkswagen defeat devices.

LolaDrives [3,4] is an Android application which has been developed to monitor RDE tests, and to provide the driver with run-time verification and information about the test drive. The application connects the mobile phone through Bluetooth to a publicly available and affordable dongle that connects to the on-board diagnostics (OBD) port of the vehicle. This application is freely available, allowing the public to carry out vehicle tests against the RDE standard. However, when doing so, it is challenging to keep track of the range of RDE constraints during the driving effort, and as such, is difficult to produce driving cycles that are a valid RDE test. In this work, we present LolaPrompts, which augments LolaDrives with the ability to provide the drivers with audible prompts and explanations in order to facilitate the production of valid RDE tests. For the sake of reproducibility and transparency, the tool developed and used in our study and the experiment data are all available online [14]. It also contains a video detailing the app and demos of the app.

In this paper, we report on the design of LolaPrompts and its empirical evaluation in real driving emission tests with volunteer participants. [4] Through our empirical evaluation of LolaPrompts, we discuss the research questions below:

(**RQ1**) Does the prompt system in LolaPrompts increase the ease to perform valid RDE tests?
(**RQ2**) What additional improvements do users suggest be made to improve the system further?

To answer these questions, we ask the drivers to drive an RDE test both using our system, and without, in order to evaluate the effectiveness of LolaPrompts; after the test, we interview the drivers to receive feedback for further improvements.

To summarise, the contributions of this paper are as follows:

1. We provide an augmented version of LolaDrives, in order to guide the driver on top of the verification system.
2. We show that our modifications to LolaDrives help achieve more valid RDE tests compared to those without the application.
3. We provide the data from our user study, as well as additional extensions to LolaDrives, implementing the feedback received, to further improve the usability of the app.

2 Related work

Real Driving Emissions are expected to vary in on-road conditions, for the same vehicle model, in different conditions [10] [13]. Millions of diesel-powered cars have been fraudulently equipped with tampered emission cleaning systems [15].

A single Nitrogen Oxide cheating device in a diesel car out of 1000 cars, can contribute to about 2% increased infant mortality rate [11]. The detrimental effects of these emissions on the environment and public health are evident and as such, the development of methods to better test vehicles and reveal any cheating device or under-engineered vehicle functions is crucial.

This concern can be addressed by the development of a technology which reintroduces transparency to the system. One way to improve the system's transparency is by providing runtime verification for the state of the RDE constraints

[4] Ethics review process has been followed and approval has been obtained from King's College London prior to the user studies with KCL Ethics Reference LRU/DP-23/24-41035.

and emissions of the vehicle, independently of the car's internal monitor. In this way, we enable the public, as well as policy-makers, to detect cheating devices in vehicles by analysing systems' behaviour. With the integration of autonomous systems in vehicles, it is important to recognise the ability to cheat the testing systems and facilitate and assist public scrutiny.

LolaDrives utilises RTLola [8], a stream-based runtime monitoring system extending LOLA [4]. Köhl et al. [12] present a formal description of RDE constraints and their violation, by initially extending the LOLA specification language [5] and later using only an enhanced official version of RTLola [4]. The formalisation of RDE constraints is used to determine the information that is presented to the user by LolaDrives.

The existing LolaDrives app provides the user with a visual representation of the NOx emissions, the proportions for the different driving modes and their corresponding dynamics. At the end of a test, the app will provide details on whether the RDE test that has been completed was valid [3]. The app does not currently provide the user with information on the additional constraints, or with suggestions that aim to help the user to complete a valid RDE test. We aim to address this gap in the LolaPrompts app. To this end, we develop our runtime monitoring module that in parallel with the validation using RTLola analyses the stream of data and generates visual and voice-based prompts.

Text to speech is being used in web contents to improve accessibility of applications. According to the W3C recommendations for web accessibility, audio may be used as a text alternative [1]. Incorporating a text-to-speech corpus provides the driver with information about the status of the test in an alternative form, reducing the frequency the driver needs to look at the screen in order to understand the state of the test.

This paper is a continuation of our long-standing effort to detect and characterise software and systems doping [6,2,7]. Through facilitating valid RDE tests, we can gather useful data that can be used through our doping detection pipeline and allow for a more accurate scrutiny of vehicle emission behaviour by the general public.

3 Tool and Methodology

We propose LolaPrompts as an extension to LolaDrives, where we build an extra layer of runtime monitoring on top of RTLola in order to generate and play a voice feedback mechanism. To provide the user with the most relevant instructions, we have created three main components: a velocity profiler, a trajectory analyser and a prompt generator.

The velocity profiler logs the vehicle's velocity regularly through interaction with RTLola. Using a similar runtime monitoring mechanism, it builds a profile according to the speed classifications of driving styles as the RDE constraints define them. This includes stopping time when the car is at 0 km/h, a high-speed for velocity over 100 km/h, and a very high-speed for at least 145 km/h and up to 160 km/h, which is the highest allowed limit.

This data is used in the trajectory analyser, which analyses the current state of the test, and determines the latest state and violations of each of the RDE constraints accordingly. For constraints which can make the RDE test invalid, the analyser will calculate whether and how those could be remedied, considering the maximum time left. Accordingly, the analyser determines the next driving mode the driver should aim for, and provides the user with the speed they should drive in, and the time required for this driving mode to be sufficient.

The prompt handler prioritises the prompts based on their urgency, and on the current driving style, to provide the driver with an instruction that they can currently take into consideration. The prompt is then shown to the user as well as spoken to the driver. Currently, the following constraints are included in the prompts:

1. Stopping Percentage - the vehicle must be stopped for between 6% and 30% of the urban time.
2. Average Urban Speed - the average speed must be between 15 km/h and 45 km/h.
3. High driving speed (>100km/h) - must be driven for at least 5 minutes of the test.
4. Very-high speed (>145km/h) - must account for no more than 3% of motorway time.
5. Urban proportion - the urban distance must be between 29% and 44% of the total distance of the test.
6. Rural proportion - the rural distance must be between 23% and 43% of the total distance of the test.
7. Motorway proportion - the motorway distance must be between 23% and 43% of the total distance of the test.
8. Urban Dynamics - the 95th percentile of the vehicle's jerk is below a limit based on urban average speed, and the relative positive acceleration is above a limit based on the average speed.
9. Rural Dynamics - the 95th percentile of the vehicle's jerk is below a limit based on rural average speed, and the relative positive acceleration is above a limit based on the average speed.
10. Motorway Dynamics - the 95th percentile of the vehicle's jerk is below a limit based on motorway average speed, if it is under 94 km/h else it will be 0.025 m^2/s^3, and the relative positive acceleration is above a limit based on the average speed.

The last three dynamics are too technical for an unfamiliar driver, and we only produce general prompts about the smoothness/aggressiveness of the journey if they are being violated.

The text to speech Android package [5] is used to allow for a vocal alternative to the text. The speech is evoked on the main suggestion to improve safety and accessibility by reducing the user's need to look at the screen for instructions. To avoid unnecessary reiterations of suggestions, and allow the driver to focus

[5] `https://developer.android.com/reference/android/speech/tts/TextToSpeech`

on the road, a suggestion is spoken out only in certain conditions, such as when the text changes as well as the type of prompt.

We use the RTLola specification to provide the driver with a reason for why an RDE test is invalid, when that is the case. In addition, after the completion of the test, we run a final validation through RTLola to make sure that the results are indeed a valid RDE test. After the RDE test has been running for 90 minutes, a prompt indicating whether the overall test is valid or not will appear. If the test is invalid, the prompt will also provide a clear reason for its invalidity. This will interpret the value for the invalid RDE(number), which is returned from the RTLola specification, as an indicator for the appropriate prompt. For example, “The RDE test is invalid currently because the stopping percentage of 2% is too low". This provides a specific reason for the driver as to why the test might be invalid at the current state and providing actionable guidance to help adjust the test condition towards validity.

3.1 Testing

To improve the efficiency of the testing process of the app, and the reliability of the software, we have introduced Android tests of the existing fragments in the app, and Unit tests for the events and the newly added functionality.

The initial test drives have been conducted within the development team, to evaluate the usability of the app and make design decisions, such as the ideal frequency of instructions to the driver that will provide useful information without distracting them from driving.

To perform a meaningful evaluation, the car model was not known in advance of each test drive, in line with the original user experience evaluation of the LolaDrives app [3].

As opposed to the original version of the LolaDrives app [3], the updated version, which is used in LolaPrompts, allows conducting a test even if the NOx sensor provides only raw data, and no diagnosis. This increases the range of cars that can be used, and their emissions can still be recorded.

From the test drives, some improvements have been made to the design, such as providing the driver with the reason as to why the test has failed, and an explanation which includes the relevant data. We hope that providing explanations will increase the trust of the user in the system, and improve its reliability for test conductors, as well as allow them to attempt to recover the test in the time left until the maximal duration (30 minutes).

During simulation testing, we have found that it is useful to initially have general prompts about the overall progress, as in the early stages of the test no constraint would be exceeded. This has been set to the initial 15 minutes, or until the driver completes the minimal requirement for at least one driving style. An example of a prompt during this time could be, “You are driving at the speed of 30 km/h. You have completed 2% of the required urban driving, 10% of the required rural driving and 4% of the required of the required motorway driving." When testing the app in a real-world scenario, we found that 15 minutes is too long, the drivers were doubting whether something has gone wrong. The first

10 minutes of the test is the time which has been found to provide only generic information. We have also modified the initial prompt to providing them with regular but infrequent analysis of the current progression for each driving style, rather than no information at all.

Some other modifications have been made to the instructions to improve the user experience. Rather than giving an instruction to the driver to change their driving mode, the app now informs them that they have completed sufficient driving in a certain driving style. This has been found to be more beneficial, as the ability to change the driving style is often dependent on the route, and was not as helpful a recommendation. In addition to the prompt type, the time duration has also been added as a consideration for when to speak a prompt.

4 Evaluation and User Feedback

4.1 Experimental Setup

To provide a more comprehensive evaluation of the system, twelve of test drives have been conducted with six volunteer drivers. Each driver participate in two test-drives: one without using the app (called the control group) and one with the app (called the experiment group). At the experimentation stage, the model of the vehicle was not known in advance, hence it was difficult to know the available sensors in the car that the tool can get data from. As the study is testing the effectiveness of the app in helping with facilitating valid RDE tests, we adapted the app to broaden the accepting criteria to vehicle without NOx sensors. To do this, we have loosened the criteria to allowing the tests to run even without NOx data, making the app more flexible to a wider range of different cars.

Before the test, the participants were provided with general information about the project, as well as a video about RDE tests and the app to watch. They were then required to fill in a short questionnaire, to ensure their basic understanding of the RDE test constraints and how the test would be carried out. The volunteers were expected to meet at an agreed location (somewhere feasible and close to the motorway) and were given a preplanned route to accommodate for a sufficient amount of rural and motorway driving, which they were free to adjust during the drive, (according to the app's prompts, if they were not in the control group). All participants were given a simple introduction to RDE tests and the motive behind the project, and were asked to raise any questions they might have ahead of the test drive and the follow-up questionnaire.

Each test was conducted in sets of EC or CE forms (C = control, E = experiment) [9]; we performed four sets of EC tests with participants A, B, C and E and two sets of CE tests with Participants D and F. This helps us mitigate the bias introduced by the order of tests and the learning experience of RDE test from the first test brought into the second one. Particularly, performing more EC than CE tests mitigates any bias in favour of our intervention E. Each participant was allocated the order of executing the two test drives, where the control included driving with a simple assistance, such as a timer and a map, and the experiment setup used the LolaPrompts app. After the drive, each participant was interviewed in a semi-structured interview in the form of a questionnaire. The participants were asked about their experience in both drives, and for general feedback about the app.

4.2 Results

In this section, we analyse the performance and validity of the app, using the data collected from the test drives, by comparing the results from both control and experiment groups to determine the effectiveness of the features added to the app. Table 1 provides an overview of the results, which are explained further below.

Table 1. Comparison of RDE test validity between Control and Experiment drives

Test No.	Ordering (CE or EC)	Control		Experiment	
		Valid	Constraints Broken	Valid	Constraints Broken
A	CE	No	8	No	8
B	CE	No	8	Yes	N/A
C	CE	No	5 8	No	8
D	EC	No	8	Yes	N/A
E	CE	No	7 5 8	No	9
F	EC	No	5 8 9 10	Yes	N/A

The control tests carried out were all invalid. The primary constraints commonly violated were related to the dynamic lower boundary for urban and rural driving modes, and the proportion of urban driving mode. It is a challenging constraint and in our experience smooth driving style or enabling Eco mode in the vehicle, which bound acceleration and jerk, often leads to violating it.

The experiment tests show notable improvements when compared to the control group, with three tests out of the six being valid and only one constraint being broken in each of the invalid tests. This indicates that the app's enhancements helped the drivers navigate the constraints more effectively.

User feedback User feedback was gathered through an semi-structured interview after each test drive. The main points of the feedback included:

1. Positive Feedback:
   (a) Prompts were useful, especially the text to speech.
   (b) The app provided good guidance for performing tests.
2. Constructive feedback:
   (a) Frequency of prompt distracting in certain portions of test.
   (b) Suggestions for additional features such as imperial and metric units (mph in addition to km/h).

(c) Recommendations for enhancing visual elements, including larger and more distinguishable bars, and the use of colours or pictures for different driving modes.
(d) Need for an earlier notification if the test is invalid.
(e) Bigger text for readability.
(f) Ensuring compatibility with various car systems.

4.3 Improvements

Following the experiments, we have improved the app according to driver's feedback from the tests:

1. **Metric/imperial toggle:** added option to switch between metric and imperial units, allowing easier interpretation of speed data and region-specific testing.
2. **Enhanced Visuals:** enlarged critical on-screen elements to improve visibility and ease of use for drivers, ensuring that essential information is easily accessible at a glance.
3. **Adjustable text-to-speech:** added feature to mute text-to-speech, in case the driver does not want to receive instructions for a part of the drive (e.g., a long stretch of motorway) and modified when text to speech is alerted.
4. **Option to Increase Text Size:** to improve readability and ensure all instructions and information is clear and easy to read while driving.
5. **Invalid Test Early Notification:** implemented an early notification system to inform the driver whether the test is invalidated, enabling them to restart and achieve a valid test run.

These improvements aimed to enhance the overall user experience and address specific concerns raised by test participants. The test run with the improved functionality in addition to testing the app with the simulator helped to correct mistakes.

5 Discussion

We found through our user studies that the voice prompts improve the user experience and enables more valid RDE tests. The user studies have significantly contributed to the enhancement of the app. For example, we enhanced the prompts for the acceleration and jerk constraints, which were difficult to achieve in the previous version of the app.

Potential threats to the soundness of this work is the small sample size and limited scenarios tested due to limited number of volunteers. We plan to unroll the app publicly after further testing. Across the many tests that the team and the volunteers have performed (both for preparing the app and during the designed experiments), we have never driven a single diesel car that would satisfy the maximum allowed NOx emission of 120 mg/km. On average, across all tests (valid and invalid), we measured 307.64 mg/km of NOx emission. This underscores the importance of our next steps in unrolling the app and allowing for public scrutiny.

Acknowledgements

Mohammad Reza Mousavi has been partially supported by the UKRI Trustworthy Autonomous Systems Node in Verifiability, Grant Award Reference EP/V026801/2, EP- SRC project on Verified Simulation for Large Quantum Systems (VSL-Q), grant reference EP/Y005244/1 and the EPSRC project on Robust and Reliable Quantum Comput- ing (RoaRQ), Investigation 009 Model-based monitoring and calibration of quantum computations (ModeMCQ), grant reference EP/W032635/1.

References

1. Accessibility Guidelines Working Group: Web content accessibility guidelines (WCAG) 2.1, `https://www.w3.org/TR/WCAG21/#dfn-media-alternative-for-text`
2. Biewer, S., Dimitrova, R., Fries, M., Gazda, M., Heinze, T., Hermanns, H., Mousavi, M.R.: Conformance relations and hyperproperties for doping detection in time and space. Log. Methods Comput. Sci. **18**(1) (2022). `https://doi.org/10.46298/LMCS-18(1:14)2022`, `https://doi.org/10.46298/lmcs-18(1:14)2022`
3. Biewer, S., Finkbeiner, B., Hermanns, H., Köhl, M.A., Schnitzer, Y., Schwenger, M.: Rtlola on board: Testing real driving emissions on your phone. In: International Conference on Tools and Algorithms for the Construction and Analysis of Systems. pp. 365–372. Springer (2021)
4. Biewer, S., Finkbeiner, B., Hermanns, H., Köhl, M.A., Schnitzer, Y., Schwenger, M.: On the road with rtlola. Int. J. Softw. Tools Technol. Transf. **25**(2), 205–218 (2023). `https://doi.org/10.1007/S10009-022-00689-5`, `https://doi.org/10.1007/s10009-022-00689-5`
5. D'Angelo, B., Sankaranarayanan, S., Sánchez, C., Robinson, W., Finkbeiner, B., Sipma, H.B., Mehrotra, S., Manna, Z.: LOLA: runtime monitoring of synchronous systems. In: 12th International Symposium on Temporal Representation and Reasoning (TIME 2005), 23-25 June 2005, Burlington, Vermont, USA. pp. 166–174. IEEE Computer Society (2005). `https://doi.org/10.1109/TIME.2005.26`, `https://doi.org/10.1109/TIME.2005.26`
6. D'Argenio, P.R., Barthe, G., Biewer, S., Finkbeiner, B., Hermanns, H.: Is your software on dope? - formal analysis of surreptitiously "enhanced" programs. In: Yang, H. (ed.) Programming Languages and Systems - 26th European Symposium on Programming, ESOP 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings. Lecture Notes in Computer Science, vol. 10201, pp. 83–110. Springer (2017). `https://doi.org/10.1007/978-3-662-54434-1_4`, `https://doi.org/10.1007/978-3-662-54434-1_4`
7. Dimitrova, R., Gazda, M., Mousavi, M.R., Biewer, S., Hermanns, H.: Conformance-based doping detection for cyber-physical systems. In: Gotsman, A., Sokolova, A. (eds.) Formal Techniques for Distributed Objects, Components, and Systems - 40th IFIP WG 6.1 International Conference, FORTE 2020, Held as Part of the 15th International Federated Conference on Distributed Computing Techniques, DisCoTec 2020, Valletta, Malta, June 15-19, 2020, Proceedings. Lecture Notes in Computer Science, vol. 12136, pp. 59–77. Springer (2020). `https://doi.org/10.1007/978-3-030-50086-3_4`, `https://doi.org/10.1007/978-3-030-50086-3_4`

8. Faymonville, P., Finkbeiner, B., Schwenger, M., Torfah, H.: Real-time stream-based monitoring. arXiv preprint arXiv:1711.03829 (2017)
9. Gou, M.S., Lakatos, G., Holthaus, P., Robins, B., Moros, S., Wood, L., Araujo, H., deGraft Hanson, C., Mousavi, M.R., Amirabdollahian, F.: Kaspar explains: The effect of causal explanations on visual perspective taking skills in children with autism spectrum disorder. In: 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). pp. 1407–1412. IEEE (2023)
10. Joint Research Centre (European Commission), Zardini, A., Bonnel, P.: Real Driving Emissions Regulation: European methodology to fine tune the EU real driving emissions data evaluation method. Publications Office of the European Union, `https://data.europa.eu/doi/10.2760/176284`
11. Jonson, J.E., Borken-Kleefeld, J., Simpson, D., Nyíri, A., Posch, M., Heyes, C.: Impact of excess nox emissions from diesel cars on air quality, public health and eutrophication in europe. Environmental Research Letters **12**(9), 094017 (2017)
12. Köhl, M.A., Hermanns, H., Biewer, S.: Efficient monitoring of real driving emissions. In: Runtime Verification: 18th International Conference, RV 2018, Limassol, Cyprus, November 10–13, 2018, Proceedings 18. pp. 299–315. Springer (2018)
13. Kurtyka, K., Pielecha, J.: The evaluation of exhaust emission in rde tests including dynamic driving conditions. Transportation Research Procedia **40**, 338–345 (2019)
14. Navaratnarajah, M., Armony, M., Biewer, S., Hermanns, H., Mousavi, M.R.: Lolaprompts (Feb 2025). `https://doi.org/10.6084/m9.figshare.28431935`, `https://figshare.com/articles/dataset/_/28431935/0`
15. Oldenkamp, R., van Zelm, R., Huijbregts, M.A.: Valuing the human health damage caused by the fraud of volkswagen. Environmental Pollution **212**, 121–127 (2016)
16. Union, E.: Commission regulation (eu) 2018/1832 of 5 november 2018 amending directive 2007/46/ec of the european parliament and of the council, commission regulation (ec) no 692/2008 and commission regulation (eu) 2017/1151 for the purpose of improving the emission type approval tests and procedures for light passenger and commercial vehicles, including those for in-service conformity and real-driving emissions and introducing devices for monitoring the consumption of fuel and electric energy. Official J. European Union L **301**, 1–314 (2017)