Vehicle identification number – anatomy of error occurrence

Roman Rak
Criminalistics and Forensic Science Department, University of Finance and Administration, Estonska 500, 101 00 Prague 10, Czech Republic
E-mail: roman.rak@irisident.cz

Abstract. Each vehicle manufactured after 1986 must be identified with a unique VIN (Vehicle Identification Number) by the manufacturer. This is so called individual vehicle identification, which is normally used as a primary data information key in various computer registers and vehicle information systems – in national vehicle registers, in police information systems (both national and international databases of stolen vehicles, etc.), in insurance company databases, leasing companies, in the eCALL technology etc. A VIN identifier error means that the vehicle cannot be found and it looks like everything is OK. The error rate of national and international registers is in the range of 5 % to 20 %. Most VIN errors occur in manual processes during transcribing or copying VINs from documents to information systems. This paper deals with the anatomy of VIN error occurrence in various processes and offers possibilities to eliminate them and improve the quality of the information systems that use VIN identifiers.

1. Introduction

1.1. Research background and significance of the subject
In this paper, we deal with so called individual errors [1], i.e., errors which are caused by persons who are transcribing the information from the vehicle documents to the information systems or directly from the vehicle to the documents or the information systems. These are subjective errors caused by fatigue, distraction, stress, using a PC keyboard or made deliberately [2]. The aim is to find the anatomy of the occurrence of these error types and to find the method of their general elimination.

This paper does not include objective errors, such as small letters of identifiers stamped on vehicles, corroded identifier surfaces, which make the characters illegible, etc. These errors are called “systematic” and they can be eliminated using systematic measures.

1.2. Analytic approach and examination data sources
Over 3,5 million real vehicle identifiers VIN were analyzed over 4 years [3]. This data came from computer databases of national central vehicle registers in the Czech Republic and the Slovak Republic. Data from national central databases of selected insurance companies, leasing and other private companies were also used. A special software program - the Universal VIN Decoder (VINexpert [4]) was developed for this analysis. The VIN decoder was filled with VIN data structures of all vehicle models produced in 1986-2018.

Information about vehicles, including VIN, is very sensitive customer data [5]. No test data, reference samples, could be used to find real errors that were not yet known. Therefore, a specialized SW application was developed to incorporate the VIN structures of all models of the world's leading vehicle manufacturers for the production period 1986-2018. Combination options were created for the
first 9 VIN positions. Their amount is about 2.5 million. The basic identification was done on the first 9 VIN positions in first step. Consideration was given to the fact whether specific manufacturers and their models use the so-called check digit. Subsequently, dozens of working hypotheses were created as to how a mistake could arise. The hypotheses were checked using specialized written SW, analyzed and some exceptions were compared manually. A more detailed description of the procedures used is far beyond the scope of this paper.

2. Individual errors
When transferring the VIN from the vehicle or its documents into information systems, a few typical character confusions may occur, which are specific for these activities. Based on many years of experience and analyses of real vehicle registrations, the following typical confusions can be specified:

- Visual confusion
- Position confusion
- Kinetic confusion
- Phonetic confusion
- Logical thought confusion
- Computer keyboard setting confusion

3. Visual confusion
The visual confusion results from the similarity of characters used in VIN. It occurs in the human mind when reading the characters. This confusion is usually caused by the incorrect or currently unwell sight of the person who is reading the VIN, a lack of knowledge about the normal VIN structure (especially the WMI and VIS sections). A visually similar character is entered instead of the correct one. VIN illegibility may also be caused by improper lighting, the location of the VIN identifier in the vehicle (inaccessible areas, blind spots, covering VINs with certain vehicle components, lighting conditions), ageing and operational processes (corrosion, contamination), inappropriate VIN font (stamping, engraving), etc. The character confusion is similar to the other general cases when someone is reading and transcribing the text.

In practice, the frequently confused characters are U-V, H-M, O-0, 5-S, etc. See Table 1 for more examples.

4. Position confusion
The position confusion occurs when typing on a PC keyboard to enter data into an information system. A user presses an adjacent key on the keyboard instead of the correct one. These are the alphabetic as well as the numeric keys, which are usually entered using a small numeric keypad on the right side. The position confusions may be caused by a rush, distraction, insufficient experience in typing on the keyboard, small keys or large fingers (an inappropriately manufactured or selected keyboard in terms of ergonomics). With position confusion the adjacent characters on a keyboard are swapped. Position confusions are apparent from Figure 1 and Figure 2.

![Figure 1. Possible confusions of “5” for the adjacent keys.](image1)

![Figure 2. Possible confusions of “J” for the adjacent keys.](image2)
Table 1. Basic examples of VIN confusion and practical examples.

| Type of confusion | Typical error situations | Incorrect VIN | Correct WMI |
|-------------------|--------------------------|---------------|-------------|
| Visual confusion  | U-V  H-M  E-F            | JM2BJ14L201274161 | JMZBJ14L201274161 |
|                   | 0-O  5-S  W-M            |               |             |
|                   | 2–Z  6-G  G-C            | JF1GCHLJ3WG067014 | JF1GCHLJ3W6067014 |
|                   | 6-C  3-B  ….  B-8       | WAUZZZ4804N045331 | WAUZZZ4B04N045331 |
|                   | Y-V  ….                 |               |             |
| Position confusion| J-K  J-U  J-Z            | VF30A9HR8BS064344 | VF30A9HR8BS064347 |
|                   | J-H  J-M  J-L  ….  J-N  * | JV1VW70821F709644 | YV1VW70821F709644 |
| Kinetic confusion | TMB – TBM                | SNJBAA016U0314749 | SJNBAAN16U0314749 |
|                   | WF0 – WOF                | TMBGS46Y023304024 | TMBGS46Y023304024 |
|                   | WDB – WDB  ….           | JMZBJ134221473662 | JMZBJ143221473662 |
| Phonetic confusion| V-W                      | VVGGZZZ5NZ8W000253 | VVGGZZZ5NZ8W000253 |
| Logical thought confusion | WF0 – WFO         | WFO4XXGBB41K55965 | WFO4XXGBB41K55965 |
|                     | WP0 – WFO                | WOL0AHL3582003279 | WOL0AHL3582003279 |
|                     | WOL-WOL                 | WVVZZZ1FZ9V028266 | WVVZZZ1FZ9V028266 |
| Computer keyboard setting confusion | Y-Z  Z-Y             | VF1FLB1B1EZ535433 | VF1FLB1B1EY535433 |
|                   |                          | JMYBJ12P200249824 | JMZBJ12P200249824 |

* See the keyboard layout pictures. Select any key and check the adjacent keys.

5. Kinetic confusion
Kinetic confusion means confusion of the order of characters. The user wants to write two characters, e.g. “TM”, but they reverse their order, which results in “MT”. This error is common among older people or users who frequently work with PC for a long time. It is so called “overtaking keystrokes by individual fingers”, which is caused by motor habits, which go beyond our reception capability. In the case of kinetic confusion the adjacent characters in the VIN string are swapped. This is distinct from position confusion where the adjacent keys on a keyboard are swapped.

6. Phonetic confusion
Phonetic (aural) confusion results from the same diction of inconsistently spelled letters. If a person pronounces “V” without accenting that it is actually “W”, the user enters “V”. Phonetic confusion is very much dependent on the user native language.

7. Logical thought confusion
This type of confusion results from unfamiliarity with the first three VIN characters (WMI). This confusion often occurs with vehicles manufactured in Germany. This applies specifically (among others) for the following brands (Table 2):
When specifying the WMI standard on a national level, lobbying of manufacturers played a significant role in reflecting the brand name in the WMI code. The first letter “W” comes from the former name of the state – West Germany. The second and the third letter were created on the Germany level – these letters determined the brand name. According to the international ISO standard, however, the use of “O” letters was not allowed, therefore zero (“0”) was used instead. The font used for stamping VINs does not allow to distinguish between “O” and “0” characters (if the user is not familiar with the standards). Due to this, many German vehicles have an incorrect “O” letter both in their documentation and information systems. This was caused by the fact that the information system
designers are usually not familiar with the VIN standards and, especially in the past, the entry of illegal VIN characters in the information system was not checked. Similar to “O” and “0” characters, the letters “I” and “J” can also be confused (“I” letter should not be used). In other cases, confusions are experienced due to a logical similarity to the brand, e.g., for the HEKU Company, the correct string is “WMI WHU”, however, “WHE” would rather be expected (which is incorrect).

Similarly, a very common error occurs when WMI is recorded for the Volkswagen vehicles. The Volkswagen name evokes an idea that the WMI string should be VWV instead of the (correct) WVV (or WVG), where the correct logic is W for West Germany (which results from the ISO standards) and VW for Volkswagen (Figure 3). Many incorrect strings WMI = VWV for the Volkswagen brand can be found in the registers; this is a quite common and typical error (Table 1).

| Brand       | Incorrectly transcribed WMI | Correct WMI |
|-------------|----------------------------|-------------|
| Ford        | WFO                        | WF0         |
| Porsche     | WPO                        | WP0         |
| Opel        | WOL                        | W0L         |
| Bitter      | WBI                        | WBJ         |
| HEKU        | WHE                        | WHU         |
| Holzer      | WHQ                        | WH0         |
| JAV Cars    | WIC                        | WJC         |
| Iveco Magirus | WIM                      | WIM         |
| Kögler      | WKO                        | WKO         |

Figure 3. Volkswagen WMI label = “WVG”.

8. Computer keyboard setting confusion

The cause of the confusion is incorrect (forgotten) switching between national and English keyboard layouts (using Alt+Shift keys). This results in swapping Y and Z characters in the VIN. This error is especially typical for programmers (using programmers setting of keyboard), persons who intensively use computers, extended applications, etc.

9. Conclusions

The VIN identifier correctness and quality is essential for efficient utilisation of information systems in the automotive area [6]. It is critical for vehicle search results, especially when it is necessary to obtain/disprove negative information on the vehicle [7], [8] (vehicles reported as stolen, vehicles with negative indications – either scraped, obstructed by an enforcement procedure, indicated as doubled, i.e., there is another vehicle in another registration system with the same VIN, having modified odometer data [9], etc.). The required VIN identifier quality grows with the growing number of
vehicles registered in information systems and with implementation of all new information systems (a general issue of continuous development of community computerisation), and the increasing number of various crimes and fraud related to vehicles. The demands are increasing with globalisation [10] (international exchange of information), European integration and cooperation, and with periodic stagnation periods in which the illegal grey or black economy is growing [11].

VINs have previously primarily been important for administrative and security reasons, i.e., for purposes of registration offices, police and the other security forces, for forensic documentation [12]. With inevitable progress, VINs are now being implemented into rescue systems (eCALL (Emergency call) project). Here VIN quality determines the quality of information, organisational and technical readiness of rescue crews before they arrive to the scene of an accident, and solutions for the hazardous situations which the vehicle passengers who called for help are exposed to. The exchange of technical, administrative and ownership information occurs between the vehicle, the integrated rescue system, and the other co-operating systems. Based on the accident scene, its reachability, and the vehicle technical specifications, the rescue crews define the necessary means for the rescue operation within tens of seconds. In other words – the VIN quality within information systems determines whether the rescue crew obtains all the necessary and correct information from vehicle databases (like the central register of vehicles) in time.

Today, the VIN quality is also important in commercial processes. These are the information systems of insurance offices, leasing companies, vehicle producers, service shops, and assistance services. The communication intensity between various vehicle information systems also increases here. The aim is to exchange the information, improve services in general, and increase customer satisfaction as well as management comfort with growing profits.

In practice, VIN identifier quality in many information systems suffers from quality, the error rate is relatively high, and the information system extraction is low. This article outlined the causes, mechanisms and ways that errors occur when VINs are manually transcribed into information systems.

In practice, it is necessary to carefully check the VIN quality at different levels. These are mostly single or repeated training courses, the analysis of errors occurred, and determining methods on how to remedy them in co-operation with the system key and methodology users, and not least with the management.

10. Future work
In general, the data (VIN) quality in information systems (IS) can be ensured in two ways. The first way is various system and control measures (methods, SW, etc.) during manual VIN transcription into IS [13]. This means verification of entered VINs, like elimination of illegal characters (I, O, Q), incorrect VIN lengths (different from 17 characters), control of the logical VIN structure and digits (using VIN decoders). The second, more advanced way is to avoid manual entries and use automatic VIN readings from vehicles or documents for direct transfer to the information system (Figure 4). This involves methods of using barcodes or 2D QR codes (that contain the VINs), reading VINs directly from vehicle control units with the use of the OBD – On Board Diagnostic (OBD) interface. Another way is using reference databases, where the VIN correctness is guaranteed so it can be used as a reference for other information systems.

![Figure 4](image-url)

**Figure 4.** An example of a VIN with bar code in parallel for fast, comfortable, and especially error-free transfer of the VIN to information systems.

To eliminate errors in the VIN, the following steps can be implemented:
• Introduce trivial computer controls to forbidden characters in the VIN structure (the letters O, I, Q are forbidden to use) and the length of the text string that must contain exactly 17 positions. These measures must be implemented when writing a VIN to any information system;
• To convince global vehicle manufacturers to use a unique check digit mechanism defined by ISO 3779:2009 for VIN quality checking. It seems ideal to change the status of the recommendation to the obligation to introduce a check digit.
• Create, update special VIN decoders to check VIN quality when entering this item in all information systems.
• Train all users about the meaning of the VIN, its structure, and its completion policies.

References
[1] Rak R, Kolitschova P, Kerbic P 2018 Forensic and technical aspects of vehicle identification labels 11th International Scientific and Technical Conference on Automative Safety, Casta Papernicka, Slovakia, Apr 18-20
[2] Kolitschova P, Kerbic J, Rak R 2018 Aspects of Vehicle Identification Labels Forensic engineering 29 (3) 2-6
[3] Felcan M 2008 Implementation of European Union legislation and regulations on road safety standards of the Slovak Republic Security of transport on the road pp. 122-132
[4] Matouskova M, Moravcik L, Rak R, Tallo A 2015 eCall, intelligent transport system (legal, technical, informational and psychological aspects) Slovakia, Bratislava: Magnet Press Slovakia 189-215
[5] Roubal O 2019 The duality of hedonism in the ambivalent world of polarities European Journal of Science and Theology 15(1) 203-213
[6] Sarkan B, Gnap J, Caban J 2017 Investigation of Exhaust Emissions of Vehicles with the Spark Ignition Engine within Emission Control, Transbaltica: Transportation Science and Technology, Procedia Engineering 187 775-782
[7] Moravcikova J 2016 Cross-Border Exchange of Information about Traffic Offences through EUCARIS Automotive safety 2016: proceedings of the X International Science-Technical Conference pp. 227-234
[8] Skrucau T, Sarkan B, Figlus T, Synak F, Vrabel J 2017 Measuring of noise emitted by moving vehicles Dynamics of civil engineering and transport structures and wind engineering, Book Series: MATEC Web of Conferences 107
[9] Brunova M, Rak R 2019 Forensics characteristics of vehicle theft In Porada V. et all: Criminalistics. Forensics science and cyber aspects Czech Republic 1015-1023
[10] Augustin P, Odler R 2013 The mission of the police in a democratic state in the context of globalization Securitologia: czasopismo naukowe, półrocznik 18( 2) 55-64
[11] Roubal O 2017 Sociology of Branding: "Just do it" in the "No Limits" World. Communication Today 8(1) 40-52
[12] Sekyrova J, Kopencova D, Manas R 2010 Documentation methods of forensic crime scene Czech Ministry of Interior, Prague
[13] Moravcik L, Jaksiewicz M 2018 Boosting car safety in the EU Automotive safety 2018: proceedings of the XI International Science-Technical Conference