Segmentation of a Road Vehicle Vibration Signal Using Multiple Comparison Procedures between Paired Samples

László Róbert HÁRI

Abstract: Segmentation of road vehicle vibration (RVV) signals can occur by the need to analyse or synthesise vibrations obtained in passenger cars or on the stowage of vans, trucks. A general and widely used measure to quantify RVV signals is its description via power spectral density (PSD). From a given PSD a Gaussian signal can be generated in a shaker testing laboratory. However, actual RVVs tend to have a non-Gaussian and nonstationary nature, which can be modelled as a composition of different segments, each with a different length and RMS content. For simulation purposes of nonstationary vibration signals, different approaches have been introduced yet each with its unique signal segmentation approaches. The current paper proposes a signal segmentation method implemented in the time-frequency domain in order to find segments within an RVV signal, where each segment has similarity in-between and is dissimilar to neighbouring segments. For this purpose, multiple comparison tests had been utilised between the Short-time Fourier transforms (STFT) applied on given fractions of an RVV. Different countermeasures had been applied against the Type I. error inflation.

Keywords: multiple comparison test; paired *t*-test; performance analysis; segmentation; short-time Fourier transform; type I. error rate control; Wilcoxon signed rank test

1 INTRODUCTION

A general and widely used measure to quantify road vehicle vibration (RVV) signals is its description via power spectral density (PSD). These can be utilised for packaging testing purposes according to many different international standards, like the International Safe Transit Association (ISTA) or American Society for Testing and Materials (ASTM) [1-3].The test signal obtained from PSD profiles via inverse Fourier transform with a uniformly distributed random phase has a stationary and Gaussian nature over time[4]. However, this approach can often be contradictory to the nature of real RVV signals. Many researchers have addressed the limitations of the established (PSD-based) methods [5]. It has been shown for RVV signals that "*non-Gaussianity is caused by the nonstationary nature of RVVs*" [6] which can be modelled as the composition of "*distinct segments each with an arbitrary duration and RMS level, i.e. the vibrations are nonstationary*" [4]. Scholars developed different methods for nonstationary simulations, which also encompassed the decomposition of an RVV into stationary segments, thus the necessity of changepoint detections. The introduction of improved methods with the utilised segmentation approaches follows in the next section.

A hypothesis testing-based RVV segmentation method is introduced in the current paper in order to systematically assess the information contained in the time-frequency domain of a given measurement series. The current paper consists of the interpretation of paired sample *t*-test and the Wilcoxon signed rank test in order to find spectrally similar segments within the waterfall plot of the Discrete Fourier transforms (DFT) calculated per 2 [sec] of an RVV. After the multiple comparison procedures, two different methods will be introduced against Type I. error inflation - which is accountable in multiple hypothesis testing. Different agglomerative diagnostics of the yielded segments will be briefly compared with an emphasis on the *post hoc* tests.

Vibration-related studies of different subsystems of the pavement-vehicle-package dynamic system may benefit from the current approach. The article discusses previous endeavours from packaging vibration testing, where signal segmentation is a possible approach to find homogenous segments or in other cases, transient events in RVVs. Although simulations are not discussed here, the presented method might be helpful to understand the nonstationary nature of RVVs, before defining non-standard excitation profiles either in shaker testing laboratories or in numerical simulation models.

2 REVIEW OF NONSTATIONARY SIMULATIONS

Many established vibration testing standards utilise vibration tests based on averaged PSD characteristics, which can describe the average signal power density as a function of frequency, revealing dominant frequency bands and characterise the overall RMS. The high-level acceleration and short-duration events are absent, which give vehicle vibration itself a non-Gaussian distribution [1, 2, 6, 7]. On the other hand, since PSD or DFT describe the signal in frequency domain, they introduce a time-invariant description of a given signal, simulating the vibration as steady-state signal with a Gaussian distribution in the time domain [6].

The simulation of **nonstationary** vibrations gained increasing emphasis by researches. The Split-spectra decomposition [8, 9] as one of the first initiatives decomposes the signal into low and high-amplitude events. Initially, two PSDs had been calculated based on 70% of the RVV with the lowest acceleration levels and the remaining 30% with the highest acceleration levels [6]. Further studies [10, 11] used different splitting ratios or even more classes for grouping events. The approach makes it possible to carry out tests on higher acceleration levels for periods. Therefore, the simulated signal -as a whole- is nonstationary and non-Gaussian. However, these subsequent extended periods *per se* are still stationary Gaussian vibrations [2].

The **Random Gaussian Sequence Decomposition** method [4, 12] produces a nonstationary and non-Gaussian signal for simulation purposes. From a measured PSD generated random Gaussian function is multiplied by a so-called *modulation function*, where the latter itself is produced from the combination of the distribution of the

segments' duration and the RMS value distribution. For the segmentation of RVVs, a cumulative sum - bootstrap algorithm had been applied on the analytical signal, obtained by Hilbert transform in previous studies [6].

Wavelet-based Gaussian decomposition [1, 2] uses continuous wavelet transform (CWT) to decompose vehicle vibration signals into Gaussian components, with an iterative process [6]. First, an equivalent random Gaussian signal is generated from the PSD of a measured RVV. The CWT of both the original RVV and the equivalent Gaussian signal is computed. Segments are considered as Gaussian parts, where the CWT of RVV fits within the maximum and minimum envelopes of the CWT of the equivalent Gaussian signal. These Gaussian parts are extracted from the signal. Residual parts of the first iteration are considered as the new RVV. The iteration begins again with the comparison of CWTs of both the (residual) RVV and its Gaussian equivalent. The number of iterations on the first level has a predetermined upper limit. The last residual segments of the iterations are considered as the non-Gaussian part of the signal, which is approximated by the best-fitting Gaussian process. The second level of iterations is initiated to improve the Gaussian fit of formerly resulted "Gaussian" parts.

Bayesian algorithm [13] has been introduced for detecting changepoints to partition measurement series of international roughness index into homogeneous sections with respect to changes in the level and variance and-or the autocorrelation. The Box-Cox transformation had been applied to the series of measurements, prior to an AMOC algorithm used as a Bayesian detector [6].

Shock extraction method has been introduced in [14], the original signal is decomposed into a series of approximated Gaussian segments and one segment containing shocks. The segmentation is carried out via the moving crest factor (MCF) and one-tenth peak (OTP) value. MCF is often used to index transients, but it is "*not always reliable and is often not appropriate for signals that contain strong non-stationarities*" [15]. Segments are simulated as stationary Gaussian signals which are concatenated to be the total simulated signal. Fatigue life prediction may also be unrealistic, since suffered damages may vary by the sequence of high- and low- or low- and high-stress fluctuations [16].

Actual transport distributions had been decomposed into the **sum of weighted Gaussians** with an iterative process in [17], which found that three to four Gaussians were enough for the decomposition of random vibrations. A curve- and peak fitting software based on the Levenberg-Marquardt algorithm had been used for the decomposition, which software enabled the "*setting of the height and the full width at half maximum of a Gaussian*".

Different **machine learning classifiers** had been studied in [18] designed for shock detection in RVV and assessed via ROC- and PERFO curves. It was found that obtaining the classifier's optimal operation point (OOP) via synthetic dataset was inadequate; thus, the re-calibration of the classifiers had been carried out via measured RVVs. Following the re-calibration, the detection capability of the *Decision Tree, 20NN* and *Bagged Ensemble* classifiers were found to be equivalent to random guessing, but the *Gaussian SVM* obtained an area under the curve (AUC) of 0,94.

The discipline of packaging vibration testing lately turned also towards multi-axial motions, like heave, pitch and roll vibrations as in [19], in which those were studied on different transports "*to establish the nature and level of the multi-axial vibrations that exist*". It was concluded among others that the relationships between the moving RMS of the above-given degrees of freedom were not strongly correlated but could be characterised statistically as joint distributions.

3 METHODS

The independency of samples must be ensured before evaluating hypothesis tests in the current method. Therefore the autocorrelation function of the acceleration signal had been examined prior in Fig. 1. The acceleration signal is strongly auto correlated within 2 s, which has the practical consequence that the smallest time resolution of the current segmentation algorithm is limited to 2 s, and shorter segments within STFT cannot be detected. Smaller fluctuations of the *ACF*-exceeding the confidence interval-are also noticeable in higher lags, which are assumed here non-significant, as if the signal was not auto correlated above lags corresponding to 2 s.
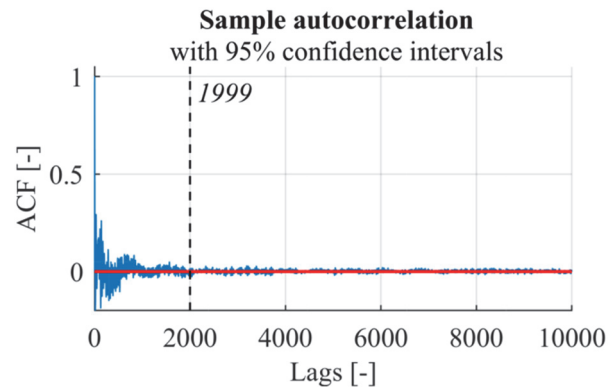

Figure 1 Autocorrelation function of the acceleration signal with blue lines, with 95 % confidence intervals indicated by the red lines and the dotted vertical line denotes lag number 1999
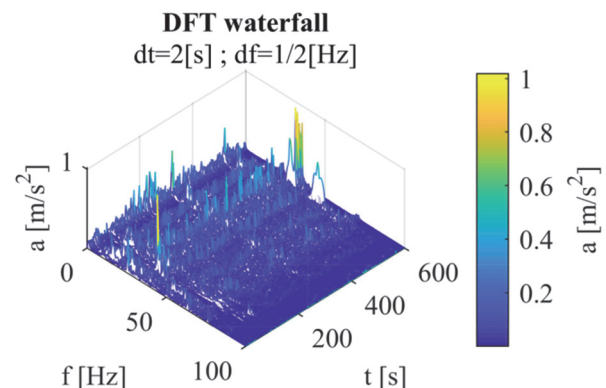

Figure 2 Waterfall plot of the discrete Fourier transforms per 2 s, where the resolution is 0.5 Hz in the frequency domain and 2 s in the time domain

The given signal in Fig. 2 represents the waterfall plot of the DFTs of an RVV signal measured on a passenger car's cockpit sampled with 1 kHz during a 600 s long journey in the study of [20]. A window frame equal to 2000 samples used for the calculation of DFTs yielded 300 timeframes.

The time-frequency domain is interpreted on an equidistant grid per axis with 0.5 Hz and 2 s resolution. The following calculations are limited to the [0; 100] Hz bandwidth. The DFT vectors have the same resolution over time; thus, it is interpretable as the paired observations of the same frequency bins over time.

The paired sample *t*-test and its nonparametric variant, the Wilcoxon signed rank test are utilised here to find the dissimilarities between consecutive DFT distributions. Two vectors of the spectrum lines will be compared on the $\alpha_0 = 0.10$ preliminary significance level. Afterwards, the resulted series of *p*-values are compared to differently adjusted significance levels.

3.1 Paired Sample *t*-test

Using a one-sample *t*-test, statistical inference can be made about the null hypothesis that the data comes from a population with a mean equal to μ. The alternative hypothesis can be formulated as that the population distribution does not have a mean equal to μ. The one-sample *t*-test is a parametric test of the location parameter when the population standard deviation is unknown.

Under the null hypothesis, the test statistic has a Student's *t*-distribution with $n - 1$ degrees of freedom [21]. The corresponding hypothesis pair for a two-sided test is, as follows:

H0: data has a mean equal μ,
HA: data do not have a mean equal μ.

The observations in each group are paired with another observation from the other group. The distribution of differences did not follow a normal distribution from the total 299 sets based on Anderson-Darling tests on 0.10 significance level. The assumption of normality of the differences is harmed in every case. However, paired *t*-test is "*robust*" to the assumption of normality [22]. The paper proceeds with the paired *t*-test despite the possible effects of the outliers, since the reason for outliers being present remains unknown, on the other hand changes in spectral peaks (on the reason of either frequency - or amplitude modulation) are in the major interest of the current examination. Alongside the nonparametric Wilcoxon signed rank test [23] is also implemented in another MCT, which is more suitable to the testing of non- normal data. The differences between each of the paired observations serve as the input to conduct a one-sample *t*-test, and the nonparametric Wilcoxon signed rank test. The paired sample *t*-test assessment of the null hypothesis, that that the data in $d = x - y$, comes from a normal distribution with mean equal to 0. The paired sample *t*-test is practically a one-sample *t*-test of the mean of differences between the paired observations, to test the null hypothesis that the pairwise differences *d* between data vectors *x* and *y* has a mean equal to 0.

H0: data in $d = x - y$ have a mean equal to 0,
HA: data in $d = x - y$ do not have a mean equal to 0.

The *p*-value is the probability of the test statistic being at least as extreme as the one observed, given that the null hypothesis is true. The choice of α is somewhat arbitrary,

although in practice values of 0.10, 0.05 and 0.01 are commonly used. The significance level will not be considered here as a part of either the acceptance or the rejection areas. It is also examined that no *p*-value falls on the significance limits in Fig. 9.

After completing STFT with 2 s window lengths resulting in Fig. 2, the current design can be considered as the paired observations of DFT values a_i in the $2i$-th secundum, in-between 0.5 Hz resolution. Let *j* denote the *j*-th difference vector and the *j*-th hypothesis pair takes the form:

$$\mathrm{H}_0^{(j)}: \boldsymbol{d}_j = \boldsymbol{a}_i - \boldsymbol{a}_{i+1} \text{ has a mean equal to 0,}$$

$$\mathrm{H}_A^{(j)}: \boldsymbol{d}_j = \boldsymbol{a}_i - \boldsymbol{a}_{i+1} \text{ do not have a mean equal to 0.}$$

When we fail to reject the $\mathrm{H}_0^{(j)}$, that would indicate that two consecutive DFT vectors have the same mean, which could indicate that these vectors are associated through their mean statistics, and those will be grouped in the same segment. In case of a significant finding, namely we reject an $\mathrm{H}_0^{(j)}$ in favour of $\mathrm{H}_A^{(j)}$ would indicate that the two consecutive DFT vectors do not have the same mean, which could indicate dissimilarity among the mean statistics of the vectors, hence they are not associated in this sense, and the initial point of a new segment can be defined.

3.2 Wilcoxon Signed Rank Test

The nonparametric Wilcoxon signed rank test is utilised here for two populations being paired. The test statistic *W* is the sum of the ranks of positive differences between the observations in the two samples, $\boldsymbol{d}_j$. The returned $p_W^{(j)}$-value corresponds to a paired, two-sided test for the null hypothesis that $\boldsymbol{d}_j$ comes from a distribution with a median equal to 0. In case the sample size is large, or the method is *approximate*, the test function calculates the *p*-value using the *z*-statistic [21].

3.3 α Adjustment

The preliminary significance limit is set to $\alpha_0 = 0.10$ in current example. Even though a high number of tests found significant *a priori*, not all of them may be considered truly significant due to α inflation. This section introduces the Type I. error inflation and the following two sections deal with two different *post hoc* procedures to control the familywise error rate (FWER).

The Type I. error rate or significance level is the probability of rejecting the null hypothesis, given that it is true [24]. If the null hypothesis is false, then it is impossible to make a Type I. error [25]. The incorrect rejection of the null hypothesis is referred to as a false positive. The probability of correctly "*accepting*" a true null hypothesis equals $S = 1 - \alpha$ and is called *specificity*.

The incorrect acceptance of the null hypothesis is called Type II. error [24], which can only occur if the null hypothesis is false, and the probability of committing this

error is called β. This second type of error is often called a false negative. The probability of correctly rejecting a false null hypothesis is $P = 1 - \beta$ and is called *power*. Types of the possible statistical inferences are followed in Tab. 1.

Table 1 Types of conclusions in a statistical hypothesis test

		Statistical inference	
		H_0 True[*]	H_0 False
Actual fact	H_0 True	Correct Decision, $S = 1 - \alpha$	Type I. error, α
	H_0 False	Type II. error, β	Correct Decision, $P = 1 - \beta$

[*]Precisely *Fails to reject* H_0

S denotes *specificity*, P denotes *power* or in other words *sensitivity*

In the simultaneous multiple comparisons of more than two groups, e.g. A, B, C the following $j = 1, \ldots, m$ hypotheses can be formulated, and the set of comparisons is referred to as a *family of test*. In the current case, $m = 299$ comparisons per MCT have been made.

Tab. 2 introduces the framework for simultaneous hypotheses testing [26], given at least two hypotheses to be tested.

Table 2 Framework of multiple comparison hypothesis tests

		Statistical inference	
		H_0 True[*]	H_0 False
	Total m	Not rejected, $m - R$	Rejected, R
Actual fact	H_0 True m_0	Correct Decisions, $m_0 - U$	Type I. errors, U
	H_0 False $m - m_0$	Type II. errors, $m - R - (m_0 - U)$	Correct Decisions, $R - U$

[*]Precisely *Fails to reject* H_0

m, m_0, R, U) … denote the number of hypotheses in the corresponding cells.

When more than one hypothesis has been simultaneously tested the probability of committing false statistical inferences increases considerably. Utilising the same Type I. error rate in an increasing number of comparisons will result in an ascending familywise error rate[1] $FWER = P(U > 0)$, i.e. it will increase the probability of at least one Type I. error during the MCT. The Bonferroni inequality can be expressed for a set of tests, as $\alpha_{fw} \leq m \cdot \alpha_{pc}$. The overall error level for the family of tests α_{fw} differs from the α_{pc} error level for one comparison. If the same α_{pc} error level is adopted for each comparison in multiple m comparisons, the overall α error level for the family of tests α_{fw} is calculated as the following procedure in Eq. (1) and Eq. (2) [27]: The probability of no α error per one comparison is $S_{pc} = 1 - \alpha_{pc}$, and the probability of no α error for the overall family of m tests is

$$S_{fw} = \left(1 - \alpha_{pc}\right) \cdot \ldots \cdot \left(1 - \alpha_{pc}\right) = \left(1 - \alpha_{pc}\right)^m \qquad (1)$$

---

[1]Other measures for accounting Type I. errors [26], like the *per-comparison error rate*, such as $PCER = E(U)/m$, where $E(\cdot)$ denotes the

for comparisons independent of each other. The family-wise Type I. error rate is:

$$\alpha_{fw} = 1 - \left(1 - \alpha_{pc}\right)^m \qquad (2)$$

Eq. (2) is also known, as α inflation and is visualised in Fig. 3 for three different preliminarily chosen typical $\alpha_0 = \alpha_{pc}$ levels.
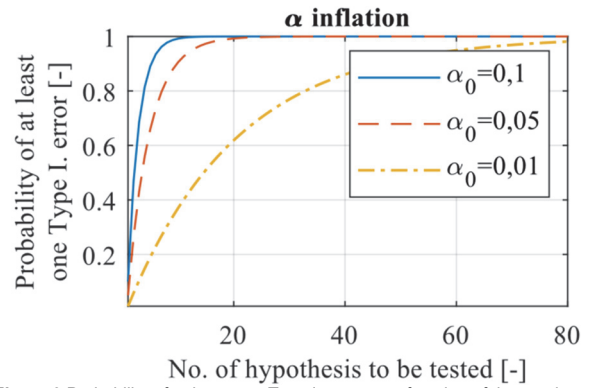
Figure 3 Probability of at least one Type I. error as a function of the number of comparisons for cases of applying the same 0.10, 0.05 and 0.01 significance levels per comparison

Different α adjustment techniques can be introduced as countermeasures to overcome the issue of the inflating likelihood of a Type I. error. The following sections discuss two of the possible processes, namely the most straightforward -sometimes referred to as too conservative-Bonferroni adjustment and its less conservative variant the Holm-Bonferroni adjustment.

Here shall be noted that for the goal of current signal segmentation method, the MCT is not utilised in the same manner as some of the typical *posthoc* tests following an ANOVA design, where the tests are interested in *all pairwise comparison* (e.g. $A \& B; A \& C; B \& C$). The number of hypotheses to be tested is relatively high; on the other hand, the design of the current experiment is only interested in the comparison of neighbouring vectors (e.g. $A \& B; B \& C$). Therefore, a consecutive setup conforms best the approach of this paper.

3.3.1 Bonferroni Adjustment

The Bonferroni method is one of the most commonly used methods to control FWER[26]. Based on the Bonferroni inequalities [28], an approach is described for "*constructing sets of simultaneous confidence intervals for the means of variables which follow a multivariate normal distribution*" [29]. The Bonferroni method is a simple method that allows many comparison statements to be made (or confidence intervals to be constructed) while maintaining an overall confidence coefficient. With an increasing number of hypotheses to be tested the Type I. error inflates, therefore the significance level is divided into the number of hypotheses tests and Type I. error is lowered in this manner. In this single-step adjustment

---

expected value; or the *false discovery rate*, such as $FDR = \{E(U/R)$, if $R > 0$; otherwise $FDR = 0$, if $R = 0\}$ are not discussed in this paper.

individual p-values $\boldsymbol{p}^{(j)}$ are compared to a new significance level $\alpha_B = \alpha_0 / m$, where α_B is the new adjusted significance level for every hypothesis, α_0 is the preliminarily chosen significance limit and m denotes the number of comparisons. The more hypotheses to be tested, the more stringent the criterion and the lower the Type I. error per comparison, but the lower the power of the test, as well.

3.3.2 Holm- Bonferroni Adjustment

Based on the Bonferroni method, Holm adjustment [30] was subsequently proposed with less conservative [31] character. Holm method, in a stepwise way, computes the significance levels depending on the p-value based rank of hypotheses. The p-values are sorted in ascending order $\boldsymbol{p}^{(s)}$ and compared against $\alpha_{HB}^{(s)}$. A step-down repeated test is performed according to ordered p-value of each comparison. As steps are progressing, comparisons are assessed with successively increasing $\alpha_{HB}^{(s)}$ adjusted error levels. It shows more power compared to the Bonferroni procedure [27]. The procedure similarly to [26] is, as follows.

The comparison stops at the first $p^{(s^*)} \geq \alpha_{HB}^{(s^*)}$, this and aftercoming hypotheses are directly declared non-significant without requiring individual comparison. In other words, let s^* be the minimal index, such that $p^{(s^*)} \geq \alpha_0 / \left(m - s^* + 1 \right)$, all the hypotheses $H_{s=1,\ldots,s^*-1}$ are declared significant and the remaining $H_{s=s^*,\ldots,m}$ hypotheses are non-significant findings.

The series of paired tests between consecutive spectrum lines resulted in $m = 299$ tests. The α inflation makes it necessary to control the FWER, for which two methods are introduced here. The following Fig. 4 and Fig. 5 represent respectively the Bonferroni single step adjustment and the Holm-Bonferroni step down procedures underneath. Because of the strong control, the α_B can become much less than the initially chosen α_0, therefore the figures are represented on a logarithmic ordinate scale.

4 RESULTS

Multiple paired t-test and Wilcoxon signed rank test of consecutive DFT vectors $\boldsymbol{a}_i$ have been assessed to find similarities in the mean tendencies of neighbouring DFT vectors of the waterfall plot. The $\boldsymbol{a}_i$ values have been considered as paired observations of the same frequency bins over time.

The initial segmentation processes showed high fragmentation of the signal, due to the high fluctuation of p-value around α_0 preliminary significance limit, as in Fig. 4. The segmentation algorithm rejects many of $H_0^{(j)}$ implying that the DFT mean tendencies do vary significantly between many of the neighbouring spectrum

vectors. Practically the number of shorter segments (in-between spectrally similar) increased.

Controlling the FWER, the Bonferroni and the Holm-Bonferroni adjustment have been applied *post hoc*. Therefore, new significance levels have been calculated, and $H_0^{(j)}$ and $H_0^{(s)}$ have been reassessed.
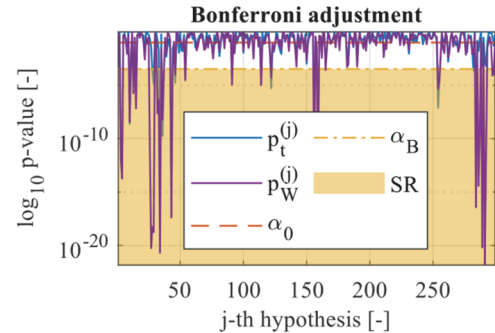

Figure 4 Multiple comparison tests with Bonferroni adjustment. $p_t^{(j)}$ and $p_W^{(j)}$ are the j-th p-value from the paired t-tests, and the Wilcoxon signed rank tests respectively; α_0 and α_B denote the *apriori* and the Bonferroni adjusted significance limits respectively and SR denotes the significance region.
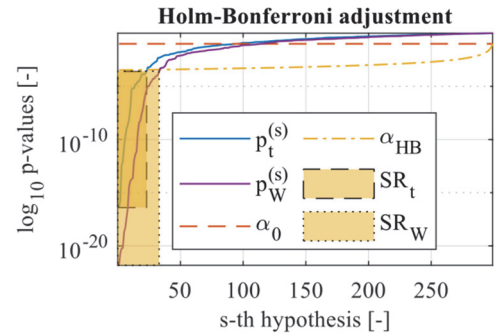

Figure 5 Multiple comparison tests with Holm-Bonferroni adjustment, with SR_t and SR_W significance regions for the t-tests and the Wilcoxon signed rank tests respectively. $p_t^{(s)}$ and $p_W^{(s)}$ are the sorted p-values from the paired t-test and the Wilcoxon signed rank test respectively; α_0 and α_{HB} denote the *a priori* and the Holm-Bonferroni adjusted significance limits respectively.
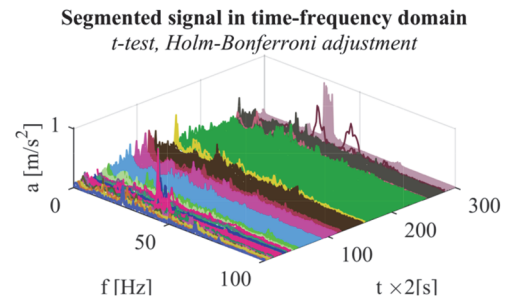

Figure 6 Exemplar of the segmentation results in the time-frequency domain. Different colours indicate different segments.
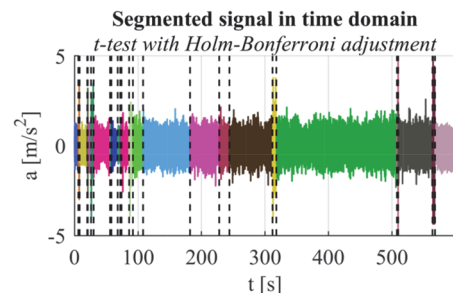

Figure 7 Exemplar of the segmentation results in the time domain. Vertical lines denote the section borders.

Since the significance limits aim to lower the FWER, i.e. Type I errors, many of *a priori* p_t - and p_W - values from different tests were considered non-significant *post hoc*. Fewer truly significant findings within the same period mean that the ratio of longer segment durations increased.

The segment lengths L and - root mean squares RMS characterise the resulting segments themselves, offering the possibility for the assessment of new aggregated statistics regarding their attributes, presented in Fig. 8. Note that the two different *post hoc* procedures gave the same solutions within either the *t*-test or the Wilcoxon signed rank test, which involves that the same hypotheses had been assessed as *significant findings* by the different α adjustments. However these identical results per *post hoc* procedures are not necessarily identical, which is further explained in the *Discussion* and in Fig. 9.

Segments lengths L tend to have still many shorter durations with only a few subseries from more extended periods in both *post hoc* procedures. The RMS distributions show a not so fragmented picture in total. The testing of the shape of distributions is not investigated here, since the 10 min duration of current measurement is not considered significant here to draw wide-term conclusions about RVVs in general.

5 DISCUSSION

The identical results of the two different α adjustment processes can be attributed to the same *p*-values in the hypotheses rejection areas per MCT, which is not a general conclusion, rather than an individual result of the current experiment. Namely, the two adjustments do not have to give necessarily the same solution (per MCT).

The most condensed explanation of this can be comprehended in Fig. 9: first, $p_t^{(s)}$ and $p_W^{(s)}$ do not have values meeting any of α_B or $\alpha_{HB}^{(s)}$. Secondly, on the exemplar of $p_t^{(s)}$, the same values are obtained above and below α_B and $\alpha_{HB}^{(s)}$, which is the case for $p_W^{(s)}$, as well. Therefore, in current dataset the result of segmentation has not changed between the *post hoc* procedures per type of MCT (*t*- or Wilcoxon- test), thus:

$$L_{t, B} - L_{t, HB} = 0 \qquad (3)$$

$$RMS_{t, B} - RMS_{t, HB} = 0 \qquad (4)$$

$$L_{W, B} - L_{W, HB} = 0 \qquad (5)$$

$$RMS_{W, B} - RMS_{W, HB} = 0 \qquad (6)$$

where the first subindices t or W denote the paired *t*-test and Wilcoxon signed rank test respectively; the second subindices B or HB denote the Bonferroni and Holm-Bonferroni post hoc procedures respectively.

However, the results of segmentation approaches differentiated by the type of MCT (*t*- or Wilcoxon- test) do differ, which is explained by the different *p*-values for the

hypothesis between the same neighbouring DFT vectors paired. Note in Fig. 8 how dotted, and solid lines differ at each colour, which is the consequence of the test statistics being different in the *t*-test and its non-parametric variant the Wilcoxon signed rank test.
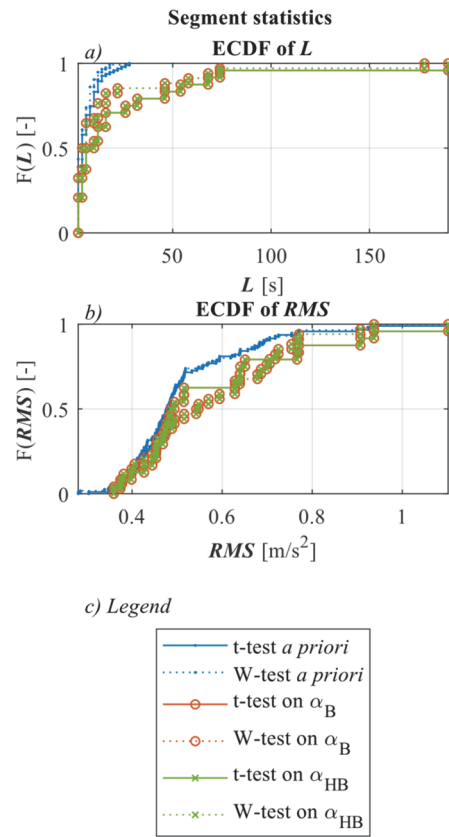
Figure 8 Empirical cumulative distribution functions of a) segment lengths, and b) segment *RMS* values. The blue lines indicate *a priori* tests, orange lines the Bonferroni *post hoc* tests overlapped by the green lines from the Holm-Bonferroni *post hoc* tests, respectively. Solid lines denote paired sample *t*-test, and dashed lines denote the Wilcoxon signed rank test.
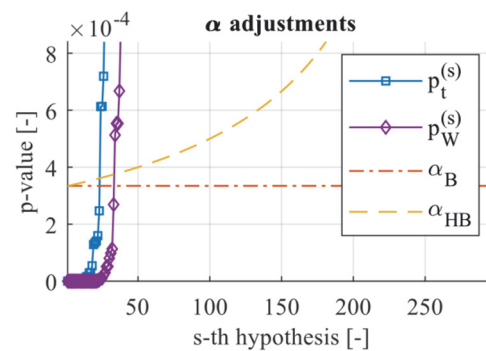Note how the green crosses overlay the orange circles.

Figure 9 Comparison of the ascendingly sorted *p*-values from the *t*-tests and the Wilcoxon signed rank tests, denoted by blue and purple lines respectively; displayed around the Bonferroni- and the Holm- Bonferroni adjusted α levels, denoted by orange and yellow lines respectively.

6 CONCLUSIONS

A multiple comparison test (MCT) procedure has been introduced in the current paper in order to find spectrally similar segments within the time-frequency domain of a road vehicle vibration (RVV) signal. The spectral values of neighbouring DFT vectors α_i have served as the inputs of paired *t*-test and its non-parametric variant the Wilcoxon

signed rank test. Two different *post hoc* procedures, namely the Bonferroni and the Holm-Bonferroni adjustments were introduced after both types of test, as countermeasures against the accountable α inflation in case of MCT.

Following studies will investigate the use of different statistical tests for MCT, such as the two-sample *t*-test and its non-parametric variant the Wilcoxon rank-sum test. However, besides the tests of central tendencies of DFT vectors, their spectral shape can be considered as well. In the latter case, the Kolmogorov-Smirnov type test seems an applicable approach.

The current article introduced and utilised two FWER-controlling procedures, but the family of FDR controlling methods holds other opportunities in *post hoc* procedures, as well. Still, the relationship between the number of *p*-values and the *post hoc* procedures must be assessed; and the sensitivity analysis of MCT shall be implemented in case of "small" amplitude and/or frequency modulations within the STFT.

The current article introduced the establishment of a novel segmentation algorithm, whereas the procedures were capable of finding spectrally similar segments within the time-frequency domain of a vibration signal. Similarities of consecutive time-slices of the STFT waterfall plot were assessed based on mean tendencies. The MCT approach to RVV segmentation purposes seems a promising solution, which the author finds worth to study further.

7 REFERENCES

[1] Griffiths, K. R., Hicks, B. J., Keogh, P. S., & Shires, D. (2016). Wavelet analysis to decompose a vibration simulation signal to improve pre-distribution testing of packaging. *Mechanical Systems and Signal Processing, 76-77*, 780-795. https://doi.org/10.1016/j.ymssp.2015.12.035

[2] Griffiths, K. R. (2012). *An Improved Method for Simulation of Vehicle Vibration Using a Journey Database and Wavelet Analysis for the Pre-Distribution Testing of Packaging*. University of Bath.

[3] Zhou, H. & Wang, Z. W. (2019). Comparison study on simulation effect of improved simulation methods for packaging random vibration test. *Packaging Technology and Science, 32*(3), 119-131. https://doi.org/10.1002/pts.2421

[4] Rouillard, V. & Sek, M. A. (2010). Synthesizing nonstationary, non-Gaussian random vibrations. *Packaging Technology and Science, 23*(8), 423-439. https://doi.org/10.1002/pts.907

[5] Kipp, W. I. (2001). Accelerated Random Vibration With Time-History Shock. *Measurement*.

[6] Lepine, J., Rouillard, V., & Sek, M. (2015). Review Paper on Road Vehicle Vibration Simulation for Packaging Testing Purposes. *Packaging Technology and Science, 28*(8), 672-682. https://doi.org/10.1002/pts.2129

[7] Zhou, H. & Wang, Z.-W. (2018). A new approach for road-vehicle vibration simulation. *Packaging Technology and Science, 31*(5), 246-260. https://doi.org/10.1002/pts.2310

[8] Chonhenchob, V., Singh, S. P., Singh, J. J., Stallings, J., & Grewal, G. (2012). Measurement and Analysis of Vehicle Vibration for Delivering Packages in Small-Sized and Medium-Sized Trucks and Automobiles. *Packaging Technology and Science, 25*(1), 31-38. https://doi.org/10.1002/pts.955

[9] Singh, J., Singh, S. P., & Joneson, E. (2006). Measurement and analysis of US truck vibration for leaf spring and air ride suspensions, and development of tests to simulate these conditions. *Packaging Technology and Science, 19*(6), 309-323. https://doi.org/10.1002/pts.732

[10] Wallin, B. (2007). Developing a Random Vibration. *IAPRI Symposium*, 42–54.

[11] Kipp, W. I. (2008). Random Vibration Testing Of Packaged-Products : Considerations For Methodology Improvement. *Packaging (Boston, Mass.)*, 1-12.

[12] Rouillard, V. (2007). The synthesis of road vehicle vibrations based on the statistical distribution of segment lengths. *5th Australasian Congress on Applied Mechanics, ACAM 2007*, 1-6.

[13] Thomas, F. (2005). Automated road segmentation using a Bayesian algorithm. *Journal of Transportation Engineering, 131*(8), 591-598. https://doi.org/10.1061/(ASCE)0733-947X(2005)131:8(591)

[14] Zhou, H. & Wang, Z. W. (2018). A New Approach for Road-Vehicle Vibration Simulation. *Packaging Technology and Science, 31*(5), 246-260. https://doi.org/10.1002/pts.2310

[15] Lepine, J., Rouillard, V., & Sek, M. (2017). On the Use of Machine Learning to Detect Shocks in Road Vehicle Vibration Signals. *Packaging Technology and Science, 30*(8), 387-398. https://doi.org/10.1002/pts.2202

[16] Raouf, A. I. (2017). *Handbook of structural life assessment*. Wayne State University, USA: John Wiley & Sons Ltd.

[17] Bonnin, A.-S., Nolot, J.-B., Huart, V., Pellot, J., Krajka, N., Odof, S., & Erre, D. (2018). Decomposition of the acceleration levels distribution of a road transport into a sum of weighted Gaussians: Application to vibration tests. *Packaging Technology and Science, 31*(8), 511-522. https://doi.org/10.1002/pts.2375

[18] Lepine, J., Rouillard, V., & Sek, M. (2019). Evaluation of machine learning algorithms for detection of road induced shocks buried in vehicle vibration signals. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 233*(4), 935-947. https://doi.org/10.1177/0954407018756201

[19] Rouillard, V. & Lamb, M. (2020). Some characteristics of the heave, pitch and roll vibrations within urban delivery routes. *Packaging Technology and Science, 33*(3), 113-121. https://doi.org/10.1002/pts.2491

[20] Irvine, T. (2004). Vibrationdata.com. *August 2004 Newsletter*.

[21] The MathWorks Inc. (2020). MATLAB 9.5.0. (R2018b). Natick, Massachusetts: The MathWorks Inc. Retrieved from https://www.mathworks.com/help/stats/ttest.html

[22] Moore, D. S. (2010). *The Basic Practice of Statistics* (Fifth.). New York: W. H. Freeman and Company.

[23] Gibbons, J. D. & Chakraborti, S. (2003). *Nonparametric Statistical Inference, Fourth Edition: Revised and Expanded* (4th ed.). New York: Marcel Dekker, Inc.

[24] Salkind, N. (2010). *Encyclopedia of Research Design*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. https://doi.org/10.4135/9781412961288

[25] Lane, D. M. *Introduction to Statistics* (Online Edi.). (1) Rice University, (2) University of Houston, Downtown Campus. Retrieved from http://onlinestatbook.com/

[26] Chen, S. Y., Feng, Z., & Yi, X. (2017). A general introduction to adjustment for multiple comparisons. *Journal of Thoracic Disease, 9*(6), 1725-1729. https://doi.org/10.21037/jtd.2017.05.34

[27] Kim, H.-Y. (2015). Statistical notes for clinical researchers: post-hoc multiple comparisons. *Restorative Dentistry & Endodontics, 40*(2), 172. https://doi.org/10.5395/rde.2015.40.2.172

[28] Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*.

[29] Dunn, O. J. (1959). Estimation of the Medians for Dependent

Variables. *The Annals of Mathematical Statistics, 30*(1), 192-197. https://doi.org/10.1214/aoms/1177706374

[30] Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70.

[31] Lee, S. & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology, 71*(5), 353-360. https://doi.org/10.4097/kja.d.18.00242

NOTATIONS

$\boldsymbol{a}_i$	/ m/s^2	DFT acceleration vector
$\boldsymbol{d}_j$	/ m/s^2	DFT acceleration differences vector
α	-	Type I. error
α_0	-	preliminary significance limit
α_{pc}	-	per-comparison Type I. error rate
α_{fw}	-	familywise Type I. error rate
α_B	-	Bonferroni adjusted significance limit
$\boldsymbol{\alpha}_{\mathrm{HB}}^{(s)}$	-	s-th sorted value of Holm-Bonferroni adjusted significance limits
p	-	p-values
$\boldsymbol{p}_{\mathrm{t}}^{(j)}$	-	j-th of sorted p-values from t-test
$\boldsymbol{p}_{\mathrm{t}}^{(s)}$	-	s-th of sorted p-values from t-test
$\boldsymbol{p}_{\mathrm{W}}^{(j)}$	-	j-th p-values from Wilcoxon signed rank test
$\boldsymbol{p}_{\mathrm{W}}^{(s)}$	-	s-th of sorted p-values from Wilcoxon signed rank test
$\boldsymbol{L}$	/ s	Segment lengths
$\boldsymbol{RMS}$	/ m/s^2	Segment root mean squares

ABBREVIATIONS

ACF	Autocorrelation function
AUC	Area under the curve
AMOC	at-most-one-change assumption
ANOVA	Analysis of variance
ASTM	American Society for Testing and Materials
CWT(s)	Continuous wavelet transform(s)
DFT(s)	Discrete Fourier transform(s)
ECDF	Empirical cumulative distribution function
FDR	False discovery rate
FWER	Familywise error rate
ISTA	International Safe Transit Association
MCF	Moving Crest factor
MCT	Multiple comparison test
OOP	Optimal operation point
OTP	One-tenth peak value
PCER	Per-comparison error rate
PERFO	Pseudo energy ratio - fall-out curve
PSD(s)	Power spectral density(ies)
RMS	Root mean square
ROC	Receiver operating characteristic
STFT	Short-time Fourier transform
SR	Significance region

Contact information:

László Róbert HÁRI, MSc
Department of Logistics and Forwarding, University of Győr,
Egyetemtér 1, 9026 Győr, Hungary
E-mail: hari.laszlo@sze.hu