



HAL
open science

Influence of structural variability upon sound perception: Usefulness of fractional factorial designs

Vincent Koehl, Etienne Parizet

► To cite this version:

Vincent Koehl, Etienne Parizet. Influence of structural variability upon sound perception: Usefulness of fractional factorial designs. *Applied Acoustics*, 2006, 67 (3), pp.249-270. 10.1016/j.apacoust.2005.06.002 . hal-00414849

HAL Id: hal-00414849

<https://hal.science/hal-00414849>

Submitted on 26 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Influence of structural variability upon sound perception: usefulness of fractional factorial designs

Vincent Koehl^{*}, Etienne Parizet

*Laboratoire Vibrations Acoustique, Institut National des Sciences Appliquées,
F-69621 Villeurbanne Cedex, France*

Abstract

The present paper introduces an efficient and time-saving approach for the evaluation of the consequences of structural uncertainties on sound perception. Its aim is to validate the use of fractional factorial designs for perceptual assessment of a model system. A test bench was used, which allowed to accurately control the variability of several structural design parameters. Sounds emitted by the bench were recorded with a dummy head and submitted to listeners during two experiments, in which they had to evaluate the dissimilarity of each sound to a reference, representing the nominal state of the device. In the first experiment, six factors, assumed to be independent, were used to define a fractional factorial design. As an analysis of variance showed that two interactions between factors should have been taken into account, a second experimental design was developed to quantify these interactions. These two experiments allowed to define an accurate model of sound perception, describing the effect of each factor on the perceived dissimilarity. Thus it was possible to relate the variability of the structure to the perception of the sound emitted with few experimental effort.

Key words: Structural uncertainties; Sound Perception; Fractional factorial designs; Taguchi tables; Listening test; Dissimilarity

1 Introduction

Because of mechanical variability affecting its structure, an object resulting from an industrial production can exhibit considerable variability in its vibratory and acoustical behavior [1-4]. For instance, Bernhard and Kompella [1] showed that, on a large panel of cars, the frequency response functions due to air-borne and structure-borne excitations could exhibit non-negligible

^{*} Corresponding author. Tel:+33-4-72-43-70-37; fax:+33-4-72-43-87-12.
Email address: vkoehl@lva.insa-lyon.fr (Vincent Koehl).

amplitude fluctuations and resonance frequencies shifts. The general problem is to determine whether structural dispersions may also give rise to perceptual dispersions. In other words, can the perception of the sound emitted by an object be significantly modified by variability affecting its structure? Even though consequences of uncertainties on the radiated sound have been studied [3,4] and can be objectively predicted, it is not yet possible to link the consequences of these uncertainties to the perceptual aspect. The aim of this paper is to present a tool to evaluate the acoustical outcomes of structural uncertainties on sound perception.

It might be assumed that the knowledge of Just Noticeable Differences (JNDs) should allow to predict that influence. JNDs are the perceptual thresholds above which a listener can perceive a variation of a sound feature. However, though the loudness' JNDs are well known (see [5] for a review on the topic), this is not the case for some other psychoacoustic metrics such as roughness and fluctuation strength. Moreover, it is not possible to predict the sound features that will be used by listeners to evaluate differences between complex sounds. The knowledge of JNDs alone does not provide information about the type of indicators that are used to differentiate stimuli. But they provide complementary information about the possible relevancy of the psychoacoustic metrics used by listeners for the differentiation task.

Therefore, the only way of measuring a small perceptual difference between complex sounds is an adequate listening test. In order to obtain statistically significant results, studies about structural uncertainties generally involve a large number of recordings. As an example, in [1], the sample group was composed of ninety-nine cars. For a perceptual study, such a large number of sounds is far too high. To evaluate the influence of relevant variability parameters with a reasonable number of sounds, efficient fractional factorial designs can be used. In such experimental designs, which are often used in many industrial applications [6,7], several factors are varied simultaneously according to a special experimental layout. The goal is to use a systematic approach for experimentation such that each experiment provides relevant information.

Fractional factorial designs have been used in a lot of studies aiming to improve processes and to spare measurement time [8,9]. However, their main application is the field of physical measurement. Up to now, published studies using fractional factorial designs for perceptual purposes involved a few number of factors [10] and disregarded any possible interaction. In [11], fractional factorial designs have been used to correlate the sound quality of a vacuum cleaner to the spectral content of sounds. If the presence of significant interactions is presumed, full factorial designs are preferred [12]. Fractional factorial designs have not yet been used so far to evaluate the consequences of structural uncertainties on sound perception, which was the main purpose of this study.

If this approach, which has already shown its efficiency for the understanding of objective vibro-acoustic data, proves to be reliable for the analysis of subjective data, it enables to spare much measurement and testing time.

The goal here was to introduce a model of subjective dissimilarity and to build a reliable predictive tool with fractional factorial designs.

2 Experimental setup for sound recordings

The listening test stimuli were sounds emitted by a device on which several structural variability parameters could be controlled.

2.1 Apparatus and experimental design

The test bench in this study was an electric machine on which dispersions, caused by typical variability of rotating machines, were simulated. Six variability parameters, considered as relevant ones, were selected initially:

- 'axial misalignment' (A) and 'angular misalignment' (C): misalignments were applied to the elastic coupling by acting on the bearings that support the drive shaft.
- 'distance between gears' (B): the distance was measured and adjusted.
- 'outer ring inclination' (D): the ring inclination was forced on the bearing shown on the diagram.
- 'dynamic unbalance' (E): unbalancing masses were mounted on a flywheel attached to the drive shaft.
- 'magnetic brake torque' (F): fluctuations of the magnetic brake torque were caused by modulating the feed current.

According to the fractional factorial designs terminology, these variability parameters were the design factors. These factors were chosen because of their easily recognizable spectral signature [13]. Moreover, they could be precisely controlled on the test bench. Three levels were assigned to each factor to characterize their influence. Table 1 summarizes the factors chosen and their levels.

Level 1 of each factor is its nominal state, levels 2 and 3 are typical values of misplacement caused by uncertainties. These variability parameters have been distributed all along the kinematic chain, as shown in Fig. 1.

The number of possible factors combinations is 729 (3^6). Hence a full factorial design for this case would have required too many measurements and too much testing time, the listener's task being by far too difficult. Therefore, a fractional factorial design was chosen. Among fractional designs, the most commonly used have been defined by Taguchi and Konishi [14]. According to Taguchi tables, an orthogonal design was chosen to quantify the effects of each factor with a low number of measurements. Here orthogonal means that each factor level is equally combined with other factor levels. Assuming that there was no interaction between factors, the experiment table referenced as L_{18} in [15], needing only 18 measurements, was used.

Table 2 shows the combinations of factor levels for each measurement. Four additional measurements were added to the L_{18} design to check the validity of the model resulting from the results of the experimental design. The measurement referenced as $S1-1$ corresponded to the reference condition of the bench, in which each factor was at its nominal level.

2.2 Measurement procedure

First, for each measurement, the test bench was set to the configuration indicated by the corresponding line of Table 2. Each of the six factors was set to the level defined by the measurement number. Once the test bench was adjusted, its radiated noise was digitally recorded ($f_s=44100$ Hz, 16-bits resolution) using an acoustic dummy head (Brüel & Kjær type 4100) equipped with two free-field microphones (Brüel & Kjær type 4189). All recording devices were located approximately 1.5 m distant from the test bench (see Fig. 2). To be used as stimuli for the listening test, these recorded sounds were limited to 3 s and submitted to a short fade in and fade out to avoid clicks.

3 First listening test: determination of main effects

3.1 Subjects

Subjects were students and members of the laboratory. They were eight women and twenty-two men, aged from 22 to 61 (average age=31, standard deviation=10). Eight of them had already taken part in listening tests. All subjects reported normal hearing.

3.2 Test procedure

Sounds were presented for binaural hearing (using a set of Sennheiser HD600 headphones) at their original level, ranging from 75.6 to 85.5 dBA. Listeners were asked to evaluate the dissimilarity between each test sound ($S1-i$) and the reference one ($n^{\circ}S1-1$ in Table 2). Pretests were done in using paired comparisons, the reference sound being randomly presented either in first or second place. Subjects could listen to a pair of sounds as often as they wanted to. Thereafter they had to assess the dissimilarity of the sound pair on a continuous scale running from "identical" ("identique") to "very different" ("très différent"). Instructions for listeners were displayed on the screen before the test start. A necessary condition for using fractional factorial designs is that the answer is measured on a continuous scale. This experiment did not give evaluable results, as sounds were assessed to be either very close or very different from the reference. Such dichotomic answers could not be regarded as continuous.

To allow the listeners to refine their ratings, a mixed method (see Fig. 3), i.e. a combination between absolute rating and comparison, was chosen then. This test procedure was adapted from a method used for the evaluation of pleasantness [18]. All sounds (the reference, eighteen sounds from the design and four validation ones) were presented on the test window at the same time, randomly ordered (the arbitrary number attributed to each sound in the test window had no link with its number in Table 2). Sounds were rated by listeners on a continuous scale running from "identical to the reference" ("identique à la référence", dissimilarity mark 0) to "very different from the reference" ("très différent de la référence", dissimilarity mark 1), the rating resolution (less than one hundredth of the full scale) being almost continuous. Again instructions for listeners were displayed on the screen before the test start. Subjects could listen to each sound as often as they wanted to in order to give answers. To make their task easier and allow them to give finely graded answers, they were given the possibility of reorganizing sounds from the closest to the furthest distance from the reference. The test duration varied from 12 min to 21 min and was typically 16 min.

3.3 Data analysis

3.3.1 Dissimilarity ratings

Fig. 4 shows the mean dissimilarity marks, averaged over all subjects, in their 95% confidence interval, for the 22 sounds. The lower the dissimilarity score is, the more the sound had been perceived as close to the reference. Sound $S1-1$, according to Table 2, was the reference. It was almost unanimously rated

0. The narrow 95% confidence interval excluded the hypothesis of differently answering subsets within the listeners panel and confirmed that the analysis of the experimental design could be carried out with the mean scores.

3.3.2 *Test reliability*

To check the test reliability, two listeners did the test twice, after an interval on one week. The test-retest comparison on a same subject clearly showed similar answers, as an example is shown on Fig. 5. The difference between the two series of answers was small, indicating a very good reliability. The Pearson's correlation coefficient r for the test-retest answers of the two listeners was respectively 0.906** and 0.911** ($p < .01$).

3.3.3 *Continuity of the answer*

As shown in Fig. 6, small differences in the listeners' judgment behavior could be deduced from their answers.

Some of them answered in a way that was close to a ranking of sounds, the most representative case being shown in Fig. 6(a). The two extreme sounds were rated 0 and 1, the other ones being equally spaced between these two limits. On the other hand, as shown in Fig. 6(b), several listeners seemed to do the task in a different way and gave equal marks to the sounds they perceived as equally dissimilar to the reference.

Most listeners gave answers within these two schematic behaviors. Hence it was assumed that answers fulfilled the requirement to be given on a continuous scale and they were averaged over the panel. No normalization of individual answers was done before averaging, for two reasons:

- Most subjects made use of the full available scale, rating sounds from 0 to 1.
- Inter-individual differences could be due to different answering strategies, but also to differences in the way sounds were perceived, which should not be compensated.

3.3.4 *Factor effects*

To investigate the effect of the factors described in subsection 2.1 on the dissimilarity scores, an analysis of variance was carried out. The relevancy and contribution of each factor are shown in Table 3. Some factors had obviously much more influence than others. The factor 'distance between gears' (B) contributed to more than 50% of the total variance, while a non-negligible part

(about 20%) remained unexplained. On the other hand, 'angular misalignment' (C) and 'dynamic unbalance' (E) showed a very weak contribution to the observed variances of the answers.

A more detailed analysis should show the effect of the specific factor levels on the judgment. Averaging the scores of all configurations having in common one factor at a given level, the difference with the overall average was the effect of the considered factor level on the measured response. For instance the effect of factor 'axial misalignment' (A) at level 1 E_{A1} on the measured response is given by:

$$E_{A1} = \frac{M_{S1-1} + M_{S1-2} + M_{S1-3} + M_{S1-10} + M_{S1-11} + M_{S1-12}}{6} - \bar{M}_{S1} \quad (1)$$

where M_{S1-i} is the mean dissimilarity score (i.e. the distance to the reference) for sound n° $S1-i$. All 6 M_{S1-j} on the fraction line of Eq. (1) have in common that the corresponding sounds are emitted from configurations where the factor A is at the level 1 ($j = 1, 2, 3, 10, 11, 12$ according to Table 2). \bar{M}_{S1} is the mean of dissimilarity judgments ($\bar{M}_{S1} = \frac{\sum_{i=1}^{18} M_{S1-i}}{18} = 0.52$).

The effects of all factors at all levels are summarized in Table 4. The strongest effect (+0.301) is observed for factor 'distance between gears' (B) at level 2 and the weakest (-0.003) for factor 'angular misalignment' (C) at level 1.

The calculated effects enabled to compute the score of any possible combination of factors. If the factors were independent, the theoretical score of any factorial combination should be predictable using the following additive model,

$$\text{Dissimilarity score} = \bar{M}_{S1} + E_{A\dots} + E_{B\dots} + E_{C\dots} + E_{D\dots} + E_{E\dots} + E_{F\dots} \quad (2)$$

where $E_{A\dots}$, $E_{B\dots}$, \dots , $E_{F\dots}$ are the factor effects as described in Table 4, see Table 2 for the corresponding levels. As an example,

$$M_{S1-7} = \bar{M}_{S1} + E_{A3} + E_{B1} + E_{C2} + E_{D1} + E_{E3} + E_{F2} \quad (3)$$

3.4 Validation of the model

Four additional sounds (n° $S1-19$ to $S1-22$) were not included in the data evaluation. They were used at this stage to provide an estimation of the model accuracy. The theoretical scores of these configurations were computed using Eq. (2) and were compared to those given by the panel of listeners, the agreement being far from perfect (see Fig. 7).

None of the predicted scores lied within the 95% confidence interval of its corresponding measured one, and for sound n°21 the error was very large.

The validation appeared thus unsuccessful, which disproved the assumption of independence of factors. Some effects, not suspected so far, and presumably caused by interactions between factors of the design, should have been taken into account. As shown in Table 3, their influence could have been early suspected from the non-negligible residual variance (18.93%) of the design. To account for the interdependency of the factors, the interactions had to be determined and their effects had to be added to the existing linear model.

4 Two-way analysis of variance to determine interactions

All combinations of two factors were at least present once in the experimental design. An analysis of variance considering pairs of factors should help to highlight possible interactions. For this purpose, a two-way analysis of variance was carried out. Two types of procedures, enabling to leave aside the listeners' effects, could be used. The two-way Repeated Measures ANOVA [16] considers listeners as an independent variable, but can only be conducted with full factorial designs. Therefore, in this study, a two-way Mean Values ANOVA [17], averaging individual responses, was carried out. According to this procedure, all possible two-way interactions were checked. As shown in Tables 5 and 6 (bold lines), two interactions with significant effects were found, between 'axial misalignment' (A) and 'dynamic unbalance' (E) and between 'distance between gears' (B) and 'magnetic brake torque' (F). The relevancy of those interactions was indicated by the Fisher-Snedecor test variable F .

This analysis confirmed the conclusion that factor 'dynamic unbalance' (E) had a very weak contribution to the dissimilarity by itself; but it could not be neglected due to its interaction with factor 'axial misalignment' (A).

5 Second listening test: Quantification of interactions

Two first-order interactions had significant effects on the measured response, a new experiment had to be designed to measure their effects and to extend the initial additive model formulated in Eq. (2). Meanwhile, in order to simplify the experiment, factors 'angular misalignment' (C) and 'outer ring inclination' (D) were disregarded for the remaining part of the study, because they had very small influence and did not interact with the others factors. Furthermore, 'axial misalignment-dynamic unbalance' (AE) and 'distance between gears-magnetic brake torque' (BF) interactions could be aliased as factors C and D , which would mean that a part of the interaction effects was included in these factor effects. Aliasing typically occurs when a factor having a negligible effect

(say M) is examined in place of an influent factor (say N). Effects of N are then measured and aliased as M . Taking these factors into consideration would then only reduce the accuracy of the results instead of giving complementary information.

5.1 *Experimental design*

The need was then to design a new test to determine the effects of both interactions. As each factor could have 3 levels, each interaction had 9 levels; according to Taguchi's results, the table L_9 [15] was chosen to estimate their effects. Two L_9 designs were combined in one experiment, as shown in Table 7. Again, four additional sounds were recorded for later validation, completing the sample up to 22 sounds.

Lines 1 to 9 belonged to the first L_9 fractional factorial design and the lines 10 to 18 to the second one; lines 19 to 22 described the validation measurements. Factors and levels expressed in this table are described in Table 1.

5.2 *Measurement and test procedure*

According to the same methodology as the one previously exposed, sound recordings, perceptual tests and data processing were carried out to obtain effects of those factors and interactions on the listeners' answers. The procedure of the experiment was exactly the same as for the first one, from recordings to listening tests. The level of the recorded sounds ranged from 76.7 to 83.7 dBA. In the following, the two fractional factorial designs will be analyzed independently to observe the two interaction effects separately.

5.3 *Subjects*

During the first experiment, a rather small inter-individual variability was noted, which allowed to reduce the listeners panel by one half. Fifteen listeners, randomly chosen among the previous thirty ones, participated to this experiment. They were four women and eleven men, aged from 24 to 50 (average age=30, standard deviation=8). Again, some listeners performed the test twice to estimate its reliability.

5.4 Results

5.4.1 Data analysis

As shown in Fig. 8, the 95% confidence interval of the mean dissimilarity scores was narrower than the one observed during the first test. Since less variability factors were tested at once in this experiment, a better inter-individual agreement could be observed. This also improved the test reliability, which was better than the one observed during the first experiment.

Some sounds were common in both experiments:

- Sounds n°S1-1 (1st experiment), n°S2-1 and S2-10 (2nd experiment) represent the reference state of the test bench. As shown in Fig. 4 and Fig. 8, they were evaluated as almost identical to the reference sound (respectively 0.001, 0.002 and 0.001).
- Sounds n°S1-11 (1st experiment), n°S2-21 (2nd experiment) are generated by equivalent configuration of the test bench. They were also given similar ratings (0.801 and 0.780) indicating that the rating scales used by the listeners were almost equivalent in the two experiments.

Table 8 shows the results of the ANOVA on the two L_9 designs. The computed contributions were consistent with the ones given by the two-way Mean Values Anova (Tables 5 and 6), even though the influence of interaction 'axial misalignment-dynamic unbalance' (AE) was overestimated by the two-way Mean Values Anova. This analysis constituted a first validation for assumptions of interactions. In both cases, the residual part was not negligible. But it could be attributed to the other factors which had not been taken into account in each experiment.

The two highlighted interactions were of same nature. They both involved a factor having a strong influence ('axial misalignment' (A) or 'distance between gears' (B)) and another one having a minor contribution ('dynamic unbalance' (E) or 'magnetic brake torque' (F)). The effects of A and B were similar to the ones determined during the first experiment (see Table 3). The effects of E and F appeared here even weaker. This confirmed that the effects of the interactions were at least partially included in the effects of factors E and F or aliased under another third factor.

For both designs of the second experiment, the examination of mean scores gave factor and interaction effects on the measured responses at each level (Table ??).

5.4.2 Common model

Since factors belonging to both L_9 designs were varied simultaneously for the measurements n° $S2-19$ to $S2-22$, an additive model applicable to both designs had to be established in order to predict the dissimilarity scores of these sounds.

A model for the first L_9 design (sounds $S2-1$ to $S2-9$) can be expressed as

$$\text{Dissimilarity score} = \bar{M}_{S2-I} + E_{A...} + E_{E...} + E_{AE...} \quad (4)$$

where \bar{M}_{S2-I} is the average score of sounds $S2-1$ to $S2-9$ ($\bar{M}_{S2-I} = \frac{\sum_{i=10}^9 M_{S2-i}}{9} = 0.15$); $E_{A...}$, $E_{E...}$ and $E_{AE...}$ are the effects of factors A , E and their interaction AE .

A model for the second L_9 design (sounds $S2-10$ to $S2-18$) can be expressed as

$$\text{Dissimilarity score} = \bar{M}_{S2-II} + E_{B...} + E_{F...} + E_{BF...} \quad (5)$$

where \bar{M}_{S2-II} is the average score of sounds $S2-10$ to $S2-18$ ($\bar{M}_{S2-II} = \frac{\sum_{i=10}^{18} M_{S2-i}}{9} = 0.49$); $E_{B...}$, $E_{F...}$ and $E_{BF...}$ are the effects of factors B , F and their interaction BF .

Those two models can be combined and extended to the whole experiment:

$$\text{Score} = \underbrace{\bar{M}_{S2-I} + \bar{M}_{S2-II}}_{\text{constant term}} + \underbrace{E_{A...} + E_{E...} + E_{B...} + E_{F...}}_{\text{factors}} + \underbrace{E_{AE...} + E_{BF...}}_{\text{interactions}} \quad (6)$$

For sounds $S2-1$ to $S2-9$, $\bar{M}_{S2-II} + E_{B...} + E_{F...} + E_{BF...} = 0$ and Eq. (6) is thus equivalent to Eq. (4). In a similar way, for sounds $S2-10$ to $S2-18$, Eq. (6) is equivalent to Eq. (5).

That model, taking into account the information provided by both L_9 experimental designs, could be used to predict scores of the additional sounds ($S2-19$ to $S2-22$).

5.4.3 Model validation

As for the first set of answers, the dissimilarity of each additional sound was predicted from Eq. (6) and compared to the measured dissimilarity. Results were very close to each other, as can be seen in Fig. 8 (last four sounds at the right side). It was thus established that no other effect or interaction took any significant part in the responses. The model for the dissimilarity score proved to be valid for the additionally tested sounds.

5.4.4 Discussion for the presumed cause of interactions

The reason why factor 'distance between gears' (B) interacted with factor 'magnetic brake torque' (F) could be explained physically. The normal force on the tooth face was directly dependent from these two factors. The whining noise was thus very sensitive to the variability of these two parameters.

The explanation of the interaction between factor 'axial misalignment' (A) and 'dynamic unbalance' (E) could be found in the sound spectra [13]. The dynamic unbalance amplified the first harmonics of the rotational frequency, which can be noted when comparing the spectra in Fig. 9(a) (nominal state) and Fig. 9(c) (unbalanced state). On the other hand, the axial misalignment amplified even harmonics of the rotational frequency, visible on Fig. 9(b) (misaligned state). The interaction between these two parameters was due to the fact that common harmonics were amplified by both factors at the same time (see Fig. 9(d)) (misaligned and unbalanced state). The axial misalignment was always clearly perceivable by the listeners, which was not the case for the dynamic unbalance. Effects of E were mainly perceived through their contributions to effects of A .

6 Relation between subjective dissimilarities and psychoacoustic metrics

Using a forward linear regression approach to describe the subjective dissimilarities where the inputs were various sound quality metrics (SPL, loudness, sharpness, fluctuation strength, roughness, tonality, intelligibility) computed using a commercial sound-analysis software (Mechanical Testing and Simulation [MTS] Sound Quality 3.7.6), it appeared that listeners' answers were mainly guided by loudness. The correlation coefficient between the sound loudness (computed according to ISO532B standard) and the dissimilarity score was 0.94 ($F(1, 20) = 64.94^{***}$, $p < .001$). The second metric improving the model was Aures sharpness [19]. Including sharpness in the regression model increased the coefficient correlation to 0.98 ($F(2, 19) = 45.25^{***}$, $p < .001$).

The same held for results of the second experiment. Loudness was still correlated with the dissimilarity, though the correlation coefficient was smaller ($R = 0.8$, $F(1, 20) = 61.19^{***}$, $p < .001$). Adding sharpness to the model improved the correlation ($R = 0.9$, $F(2, 19) = 40.93^{***}$, $p < .001$). That smaller influence of loudness was due to the fact that loudness differences between sounds were smaller in the second experiment compared to the first one.

For the first experiment, ratios between maximum and minimum values of

loudness (expressed in sone [5]) and sharpness (expressed in acum) were respectively $\frac{51.9}{30.9} = 1.68$ and $\frac{1.75}{1.41} = 1.24$. In the second one, these values were $\frac{44.2}{32.6} = 1.35$ and $\frac{1.69}{1.38} = 1.22$. In any case, these ratios were higher than the just noticeable ones, as obtained from the literature (relative increase of 7% for loudness [5] and 10% for sharpness [20]). This confirmed the perceptual influence of these metrics.

7 Conclusion

Fractional factorial designs allowed to quantify all effects contributing to the assessment of dissimilarity between sounds that were different due to consequences of structural uncertainties of a mechanical model system . The use of this method efficiently reduced the number of sounds necessary for a perceptual study. The first experiment investigated the main factor effects. It enabled to establish the major trends of a dissimilarity model and the second experiment allowed to refine it by taking interactions into account. A reliable predictive tool could be constructed from these specifically designed listening tests.

However this approach does not give any continuous representation of factor effects. Effects are clearly characterized at each level, but the question about what could happen between two consecutive levels is still open. The prediction of sound perception for intermediate levels of the factors needs additional assumptions to correctly interpolate between the levels used in the fractional factorial design.

An other characteristic of this approach is the small number of controlled factors and clearly identified interactions. The type of design used in this study is well suited to reveal first-order interactions. On the other hand, as it appeared in that study, the analysis of a first fractional factorial design, considering independent factors, gives information about the interactions that should be taken into account. This allowed to design a second experiment in order to reveal the missing information. And so forth, successive fractional factorial designs can provide more and more information to refine the model.

In all, this method can be helpful for the evaluation of the perceptual consequences of variability parameters affecting several elements of a structure.

References

- [1] Bernhard RJ, Kompella MS. Measurement of statistical variation of structural-acoustical characteristics of automotive vehicles. SAE Paper 931272, 1993.
- [2] Resh WF. Some results concerning the effect of stochastic parameters on engine mount system behavior. SAE Paper 911054, 1991.
- [3] Rebillard E, Guyader JL. Calculation of the radiated sound from coupled plates. *Acta Acustica united with Acustica* 2000;86(2):303-312.
- [4] Ouisse M, Guyader JL. Vibration sensitive behavior of a connecting angle: case of coupled beams and plates. *Journal of Sound and Vibration* 2003;267(4):809-850.
- [5] Zwicker E, Fastl H. *Psychoacoustics: facts and models*. Berlin, Springer Verlag, 1999.
- [6] Shepler P.R., Willem E.F., Brunke E.C. Small bore diesel engine testing using the fractional factorial technique to evaluate oil control. SAE Paper 750770, 1975.
- [7] Pratt J.K., Mutlu H., Heil H., Hofmann B., Schmidt N., Segain J. Photoreceptor optimization via Taguchi methods. SAE Paper 870276, 1987.
- [8] Dyer TJ, Nolan TW, Shapton WR, Thomas RS. The analysis of frequency domain data from designed experiments. SAE Paper 951274, 1995.
- [9] Kwon KS, Lin RM. Robust finite element model updating using Taguchi method. *Journal of Sound and Vibration* 2005;280(1-2):77-99.
- [10] Barthod M. Contribution à l'étude des bruits dit de "graillonnement" dans les boites de vitesses automobiles. PhD Thesis, Ecole nationale supérieure des arts et métiers, Paris, Jul. 2004.
- [11] Ih JG, Lim DH, Shin SH, Park Y. Experimental Design and Assessment of Product Sound Quality: Application to a Vacuum Cleaner. *Noise Control Engineering Journal* 2003;51(4):244-252.
- [12] Dempsey TK, Leatherwood JD, Clevenson SA. Development of noise and vibration ride comfort criteria. *Journal of the Acoustical Society of America* 1979;65(1):124-132.
- [13] Boulenger C, Pachaud C. *Surveillance des machines par analyse des vibrations: du dépistage au diagnostic*. Paris, AFNOR Editions, 1995.
- [14] Taguchi G, Konishi S. *Taguchi methods. Orthogonal arrays and linear graphs. Tools for quality engineering*. Dearborn MI, American Supplier Institute Press, 1987.
- [15] Alexis J. *Pratique industrielle de la méthode Taguchi. Les plans d'expérience*. Paris, AFNOR Editions, 1995.

- [16] Howell DC. Statistical methods for psychology. Belmont CA, Duxbury Press, 1997.
- [17] Spiegel MR. Theory and problems of statistics. New York, McGraw-Hill Book Company, 1993.
- [18] Parizet E, Hamzaoui N, Sabatié G. Comparison of listening test methods: a case study. Acta Acustica united with Acustica 2005;91(2):356-364.
- [19] Aures W. Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale. Acustica 1985;59(2):130-141.
- [20] Von Bismarck G. Sharpness as an attribute of the timbre of steady sounds. Acustica 1974;30(3):159-172.

List of Figures

1	Diagram of the test bench.	18
2	Acoustic dummy head placed near the test bench in order to record the 18 samples according to the experiment table.	18
3	Screen shot of the test window.	19
4	Dissimilarity scores in their 95% confidence interval for the 22 tested sounds ($S1-1$ to $S1-18$ according to the L_{18} table and $S1-19$ to $S1-22$ for validation). The small 95% confidence interval shows a good agreement among the listeners.	19
5	Comparison of test-retest answers for a same listener; solid line is the first answer and dashed line the second one.	20
6	Two different types of listener's response; sounds ranked according to the increasing dissimilarity in (a) and placed at the same level if assessed as equally dissimilar from the reference (b).	20
7	Comparison between measured scores (solid line) in their 95% confidence intervals and computed ones (dash-dotted line) for four additional sounds.	21
8	Mean measured dissimilarities (solid line) in their 95 % confidence interval and recomputed scores (dashed line); sounds $S2-19$ to $S2-22$ are additional ones.	21
9	Comparison of the spectral signatures of the variability parameters. (a) represents the nominal state, a dynamic unbalance has been introduced on (b) and an axial misalignment on (c), both parameters are present on (d).	22

List of Tables

1	Description of factors and levels	22
2	List of test bench configurations for the first experimental design	23
3	Analysis of variance for the L_{18} design	24

4	Factor effects on the measured response, calculated after Eq. (1), for the first experiment	24
5	Interaction between 'distance between gears' (B) and 'magnetic brake torque' (F)	25
6	Interaction between 'axial misalignment' (A) and 'dynamic unbalance' (E)	25
7	List of test bench configurations for the second experimental design	26
8	Analysis of variance for the two L_9 designs	27

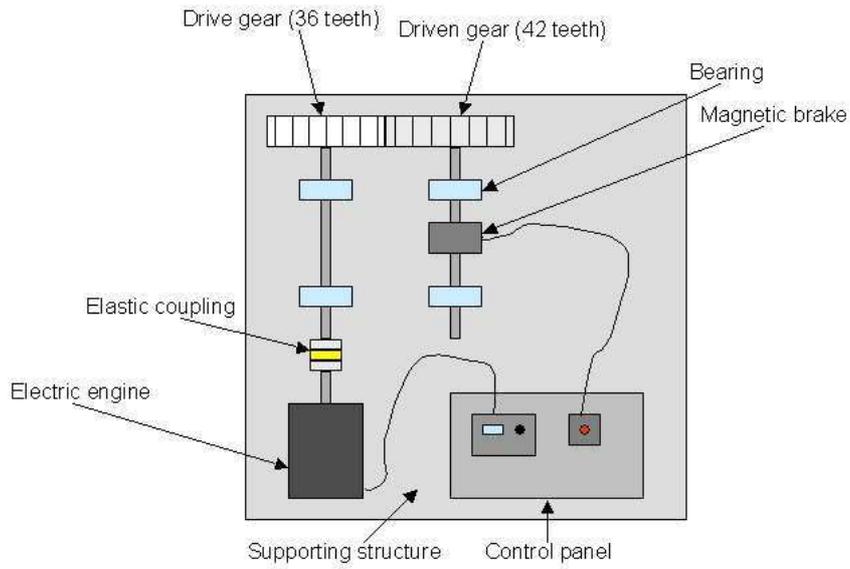


Fig. 1. Diagram of the test bench.



Fig. 2. Acoustic dummy head placed near the test bench in order to record the 18 samples according to the experiment table.

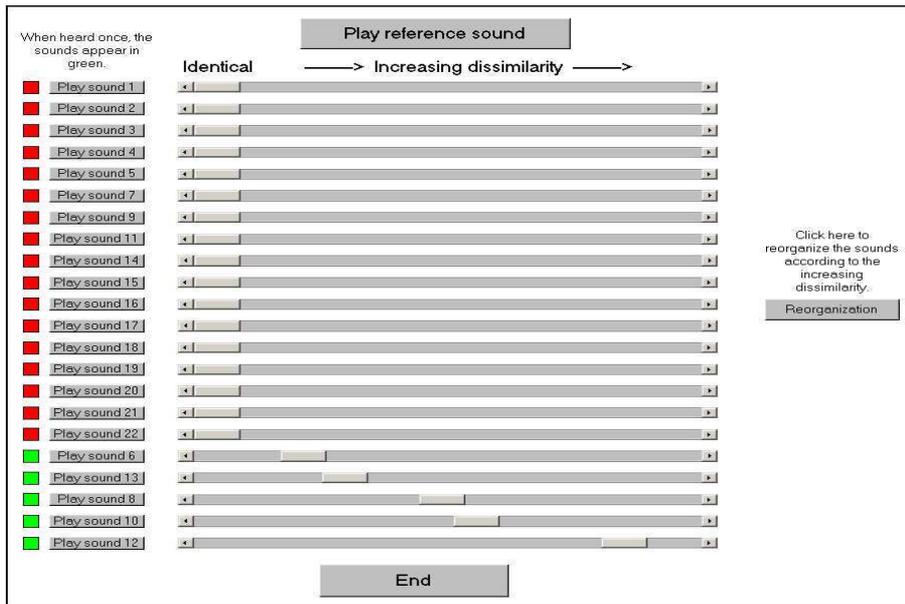


Fig. 3. Screen shot of the test window.

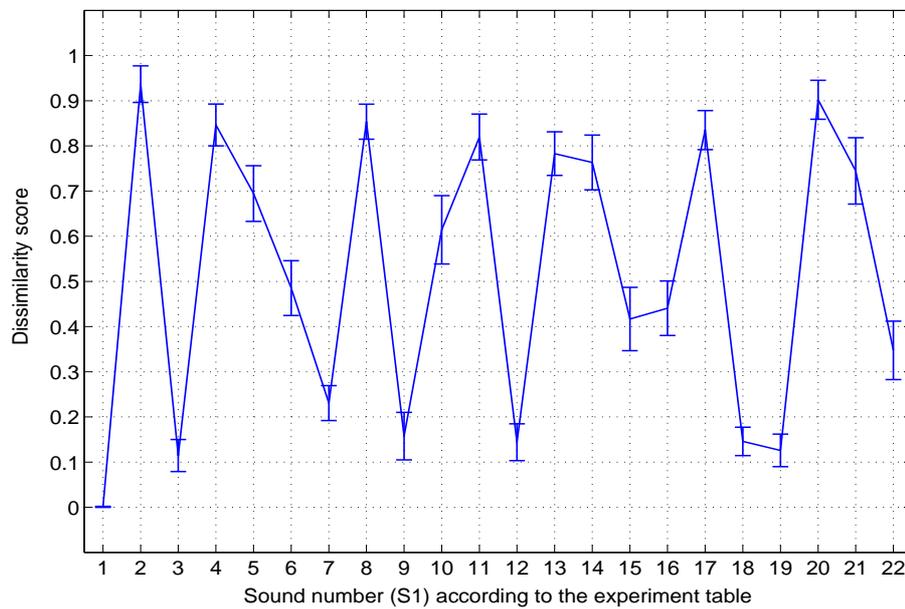


Fig. 4. Dissimilarity scores in their 95% confidence interval for the 22 tested sounds ($S1-1$ to $S1-18$ according to the L_{18} table and $S1-19$ to $S1-22$ for validation). The small 95% confidence interval shows a good agreement among the listeners.

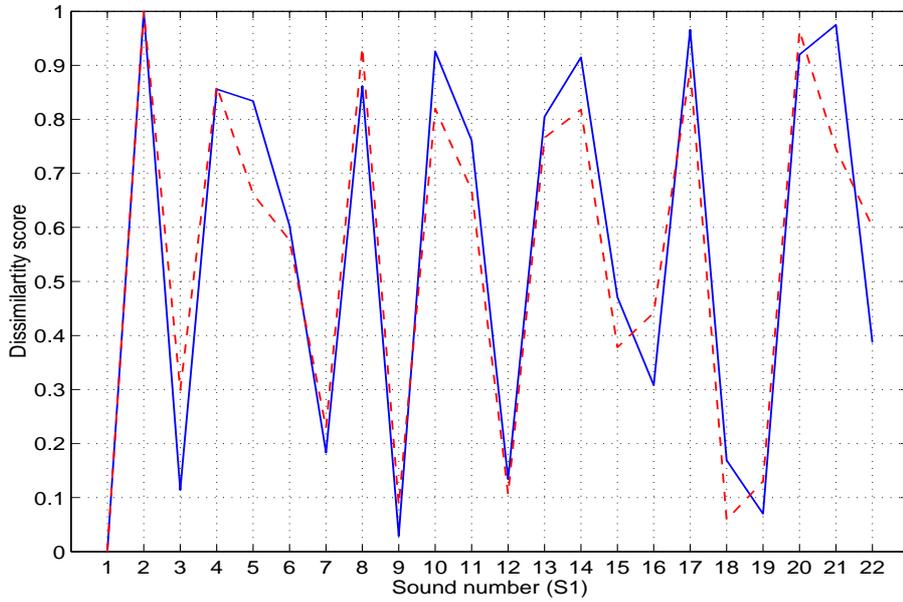


Fig. 5. Comparison of test-retest answers for a same listener; solid line is the first answer and dashed line the second one.

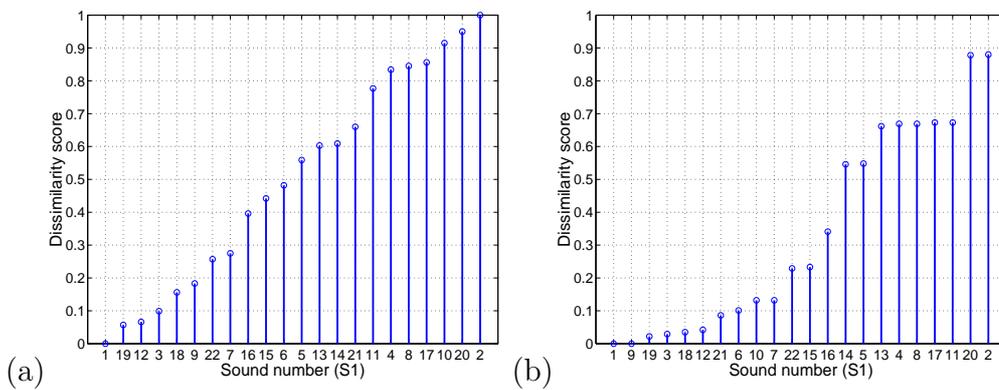


Fig. 6. Two different types of listener's response; sounds ranked according to the increasing dissimilarity in (a) and placed at the same level if assessed as equally dissimilar from the reference (b).

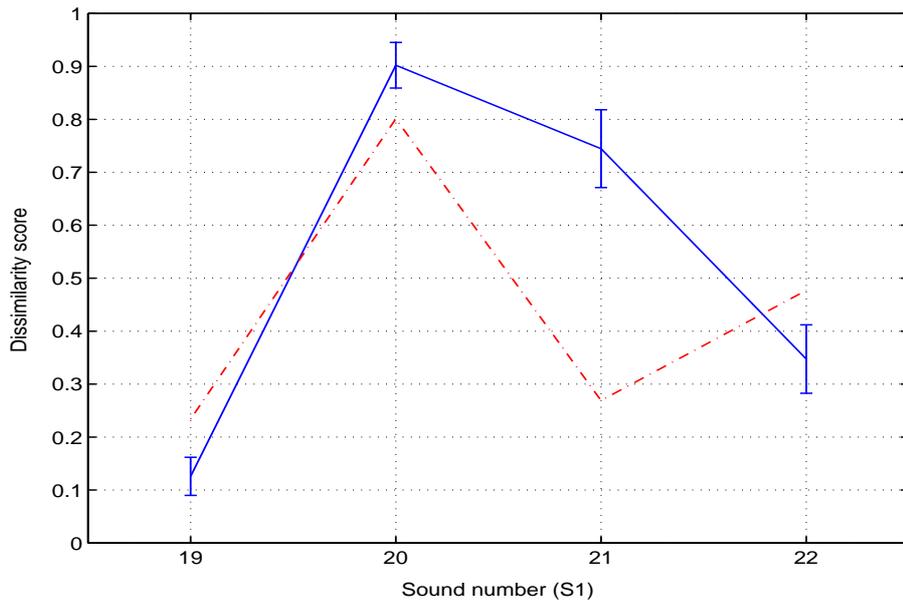


Fig. 7. Comparison between measured scores (solid line) in their 95% confidence intervals and computed ones (dash-dotted line) for four additional sounds.

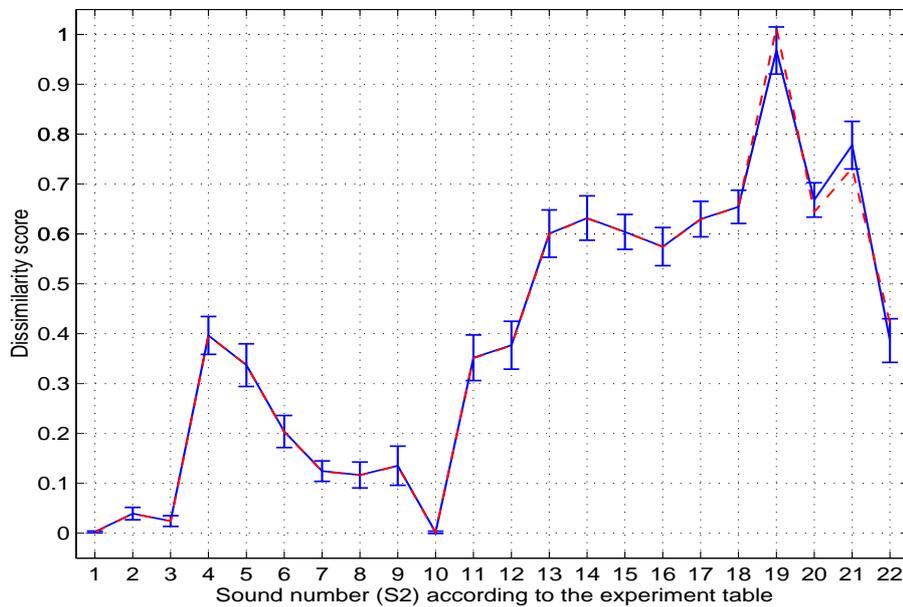


Fig. 8. Mean measured dissimilarities (solid line) in their 95 % confidence interval and recomputed scores (dashed line); sounds $S2-19$ to $S2-22$ are additional ones.

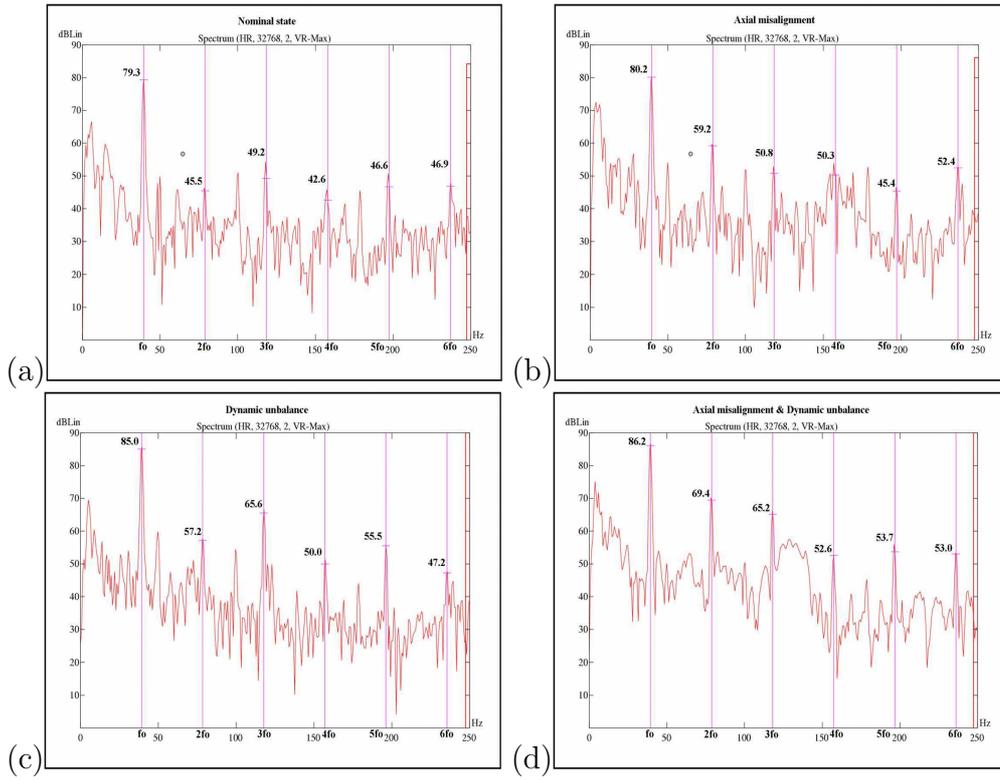


Fig. 9. Comparison of the spectral signatures of the variability parameters. (a) represents the nominal state, a dynamic unbalance has been introduced on (b) and an axial misalignment on (c), both parameters are present on (d).

Table 1

Description of factors and levels

Factor	Corresponding variability	Level and meaning		
		1	2	3
<i>A</i>	Axial misalignment	Aligned	+0.2mm	+0.4mm
<i>B</i>	Distance between gears	Aligned	-0.2mm	+0.2mm
<i>C</i>	Angular misalignment	Aligned	+0.2mm	+0.4mm
<i>D</i>	Outer ring inclination	Aligned	-1°	+1°
<i>E</i>	Dynamic unbalance	Without	$3.6 \cdot 10^{-6} \text{kg}\cdot\text{m}^2$	$7.2 \cdot 10^{-6} \text{kg}\cdot\text{m}^2$
<i>F</i>	Magnetic brake torque	2N·m	1.9N·m	2.1N·m

Table 2

List of test bench configurations for the first experimental design

Measurement number	Controlled factors					
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
<i>S1-1</i>	1	1	1	1	1	1
<i>S1-2</i>	1	2	2	2	2	2
<i>S1-3</i>	1	3	3	3	3	3
<i>S1-4</i>	2	1	1	2	2	3
<i>S1-5</i>	2	2	2	3	3	1
<i>S1-6</i>	2	3	3	1	1	2
<i>S1-7</i>	3	1	2	1	3	2
<i>S1-8</i>	3	2	3	2	1	3
<i>S1-9</i>	3	3	1	3	2	1
<i>S1-10</i>	1	1	3	3	2	2
<i>S1-11</i>	1	2	1	1	3	3
<i>S1-12</i>	1	3	2	2	1	1
<i>S1-13</i>	2	1	2	3	1	3
<i>S1-14</i>	2	2	3	1	2	1
<i>S1-15</i>	2	3	1	2	3	2
<i>S1-16</i>	3	1	3	2	3	1
<i>S1-17</i>	3	2	1	3	1	2
<i>S1-18</i>	3	3	2	1	2	3
<i>S1-19</i>	3	3	3	3	3	3
<i>S1-20</i>	1	2	3	1	2	3
<i>S1-21</i>	2	3	1	2	3	1
<i>S1-22</i>	3	1	2	3	1	2

Table 3

Analysis of variance for the L_{18} design

Variation	<i>dof</i>	Sum of squares	Mean squares	F	Contribution
<i>A</i>	2	6.01	3.00	143.02***	10.32%
<i>B</i>	2	29.81	14.90	709.36***	51.47%
<i>C</i>	2	0.29	0.14	6.81**	0.42%
<i>D</i>	2	3.64	1.82	86.60***	6.22%
<i>E</i>	2	1.40	0.70	33.24***	2.34%
<i>F</i>	2	5.99	2.99	142.50***	10.28%
Residual	527	11.07	0.02		18.93%
Total	539	58.20			

** $p < .01$; *** $p < .001$

Table 4

Factor effects on the measured response, calculated after Eq. (1), for the first experiment

Factor	Level 1	Level 2	Level 3
<i>A</i>	-0.077	0.149	-0.072
<i>B</i>	-0.030	0.301	-0.272
<i>C</i>	-0.003	-0.027	0.030
<i>D</i>	-0.108	0.091	0.017
<i>E</i>	0.001	0.062	-0.063
<i>F</i>	-0.149	0.071	0.078

Table 5

Interaction between 'distance between gears' (B) and 'magnetic brake torque' (F)

Source of variation	<i>dof</i>	Sum of squares	Mean squares	F
Distance between gears (B)	2	0.993	0.496	24.383***
Magnetic brake torque (F)	2	0.199	0.099	4.898*
Interaction (BF)	4	0.319	0.079	3.925*
Residual variation	9	0.183	0.021	
Total	17	1.696		

* $p < .05$; *** $p < .001$

Table 6

Interaction between 'axial misalignment' (A) and 'dynamic unbalance' (E)

Source of variation	<i>dof</i>	Sum of squares	Mean squares	F
Axial misalignment (A)	2	0.201	0.101	2.148*
Dynamic unbalance (E)	2	0.046	0.023	0.499
Interaction (AE)	4	1.031	0.257	5.524*
Residual variation	9	0.419	0.046	
Total	17	1.696		

* $p < .05$

Table 7

List of test bench configurations for the second experimental design

Measurement number	Contr. factors			Int.	Contr. factors			Int.
	<i>A</i>	<i>E</i>	<i>AE</i>		<i>B</i>	<i>F</i>	<i>BF</i>	
<i>S2-1</i>	1	1	1	1	1	1	1	
<i>S2-2</i>	1	2	2	1	1	1		
<i>S2-3</i>	1	3	3	1	1	1		
<i>S2-4</i>	2	1	4	1	1	1		
<i>S2-5</i>	2	2	5	1	1	1		
<i>S2-6</i>	2	3	6	1	1	1		
<i>S2-7</i>	3	1	7	1	1	1		
<i>S2-8</i>	3	2	8	1	1	1		
<i>S2-9</i>	3	3	9	1	1	1		
<i>S2-10</i>	1	1	1	1	1	1		
<i>S2-11</i>	1	1	1	1	2	2		
<i>S2-12</i>	1	1	1	1	3	3		
<i>S2-13</i>	1	1	1	2	1	4		
<i>S2-14</i>	1	1	1	2	2	5		
<i>S2-15</i>	1	1	1	2	3	6		
<i>S2-16</i>	1	1	1	3	1	7		
<i>S2-17</i>	1	1	1	3	2	8		
<i>S2-18</i>	1	1	1	3	3	9		
<i>S2-19</i>	2	1	4	3	2	8		
<i>S2-20</i>	1	3	3	2	2	5		
<i>S2-21</i>	1	2	2	3	3	9		
<i>S2-22</i>	3	1	7	1	2	2		

Table 8

Analysis of variance for the two L_9 designs

Variation	<i>dof</i>	Sum of squares	Mean squares	F	Contribution
<i>A</i>	2	1.95	0.97	147.09***	62.64%
<i>E</i>	2	0.07	0.04	5.46**	1.91%
<i>AE</i>	4	0.23	0.06	8.84***	6.72%
Residual	126	0.83	0.01		28.73%
Total	134	3.09			
<i>B</i>	2	4.17	2.08	165.22***	58.06%
<i>F</i>	2	0.67	0.33	26.52***	9.02%
<i>BF</i>	4	0.71	0.18	14.05***	9.23%
Residual	126	1.59	0.01		23.69%
Total	134	7.13			

** $p < .01$; *** $p < .001$