



HAL
open science

Design of warning sounds using an Interactive Genetic Algorithm

Jf Petiot, Sebastien Denjean, Sébastien Villa, Emmanuelle Diaz

► **To cite this version:**

Jf Petiot, Sebastien Denjean, Sébastien Villa, Emmanuelle Diaz. Design of warning sounds using an Interactive Genetic Algorithm. Forum Acusticum, Dec 2020, Lyon, France. pp.107-114, 10.48465/fa.2020.0931 . hal-03235947

HAL Id: hal-03235947

<https://hal.science/hal-03235947v1>

Submitted on 26 May 2021

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.

DESIGN OF WARNING SOUNDS USING AN INTERACTIVE GENETIC ALGORITHM

Jean-François Petiot¹ Sébastien Villa¹ Sébastien Denjean² Emmanuelle Diaz²

¹ LS2N UMR CNRS 6004, Ecole Centrale de Nantes, France

² PSA, Vélizy, France

Jean-Francois.Petiot@ec-nantes.fr

ABSTRACT

This study deals with the efficiency of Interactive Genetic Algorithms for the design of sounds. After a definition of a set of acoustic parameters of warning signals (pulse duration, interpulse interval, frequency, number of bursts, bursts interval, timbre), two methods were investigated and compared in order to understand and optimize the perceived severity level of alert sounds. The first method uses a classical D-optimal Design of experiments (DOE), the second method is based on an Interactive Genetic Algorithm (IGA). With these two methods, two within-subjects perceptual tests (30 participants) were carried out during a first session. A second session was organized with the same participants in order to assess the performances of the optimal sounds generated by both methods and to compare the results. Individual sounds and also sounds at the group level were compared, and the possible interactions between the different acoustics parameters of the alert sounds were investigated. Results show that the IGA method can be an interesting alternative to classical DOE to help the design of sounds.

1. INTRODUCTION

Sounds are nowadays used for many applications during interaction with complex products. They can be used to present information to the user (auditory display), to give a feedback on the state of a Human Machine Interface (HMI), to alert the user on potential dangers of a situation (aircraft or vehicle navigation systems [1]) or even to elicit a positive emotional state and to connote the design of a product with particular semantic dimensions (sounds of vehicles for example [2]). The design of such functional sounds is a complex design problem that necessitates coping with many constraints and dealing with perceptual aspects. In addition to the expertise of a designer, hearing tests are required in order to understand the complex relationships between acoustic parameters and perceptual dimensions.

For warning sounds in particular, knowledge of the effect of sound parameters on the perceived urgency is important to give precise recommendations [3]. Recently, new-in car technologies lead to an increasing number of sound interfaces, for example for functional sounds, navigation or Advanced Driver-Assistance

System (ADAS). Many studies propose an experimental approach with hearing tests to understand human perception of warning sounds. In [4], different hearing scenarii are proposed to study two objectives, annoyance and urgency, with a fixed experimental design. An objective measurement of warning sound urgency is proposed in [5], where the authors investigate the reaction time of participants. In addition to the perceived urgency, other dimensions such as the criticality level or the severity level of alarms can be studied [6], so as the influence of ambient noise conditions [7]. To study and understand human reaction to alarm sounds, experiments use generally a parameterized sound synthesis and classical model-based design of experiments. The limitation of such approach is that a model between the acoustic parameters and the perceptual dimension must be stated in advance, given that the exact form of the model is generally unknown.

A second category of methods, model-free in content (contrary to classical DOE, there is no model of the behavior of the respondent), but model-driven for the solution search, can be used during hearing tests. In this case, an algorithm gradually refines the propositions made to the users, for example with interactive evolutionary computation (IEC) [8]. Particular cases of IEC are Interactive Genetic Algorithms (IGA), where genetic operators such as recombination, crossover, and mutation are used to modify design samples. This method has been used for example to capture aesthetic intention of participants and to design sign sounds [9].

We propose in this paper to study the efficiency of IGA for the design of warning sounds. The first objective is to study to which extent IGA can be used to provide efficient sounds according to a given perceptual dimension. The dimension studied is the perceived severity level of alert sounds. A second objective of the paper is to compare the IGA method to the classical model-based method using a static design of experiments (DOE). For this, hearing tests based on synthesized alarm sounds were proposed to a panel of participants in a within-subject experiment. In a second experiment, the efficiency of the design solutions provided by both methods was assessed. It is important to mention that the outcomes of this study are more focused on the methodological aspects than on the design of warning sounds *per se*. The perceptual test on the rating of the

severity of alarms must be more considered as a pretext to generate consistent perceptual data than an experiment to provide realistic sounds for a dedicated application. For this reason, the purpose of the alarm sound in our study stays voluntarily vague. The only information provided to the participants is that the alarm sound is dedicated to assist the driving task of an automobile.

The paper is organized as follows. Section 2 presents a short background on interactive Genetic Algorithms (IGA) and their use in product design. Section 3 presents the material and methods for the experiments. It starts with a presentation of the sound synthesis method and the associated parameters, the description of the two experiments designed (tasks given to the participants, sound stimuli used), and the method used for the analysis of the results. Results are presented and discussed in section 4. The concluding section provides implications for sound design and perspectives.

2. BACKGROUNDS ON IGA

2.1 Principles

Genetic Algorithms (GA) are evolutionary optimization methods [10]. The principle of GA is based on iterative generations of population of individuals, converging step by step toward solutions, which are adapted to the problem. Based on the principle of Darwin's natural evolution theory, the algorithm proceeds to a selection of parents, which will spread their genetic dominant heritage in the next generation, suitable to a desired objective. Classically, the fitness evaluation of the individuals is calculated numerically with a mathematical function known beforehand. A particular category of GA, Interactive Genetic Algorithms (IGA), introduces the user in the optimization loop to assess the fitness. During each iteration, the user selects solutions (products or sounds) that he/she considers as the most interesting for the desired objective. After a number of iterations, the method may converge towards solutions that fulfill the user's objective. These algorithms are used for example to explore design spaces and to encourage creativity [11]. Since the user decides the individual fitness, there is no need for a prior and unique formulation of the fitness function. For some applications, such as exploring semantic dimensions [12] or integrating complex perceptual processes [13], this advantage is crucial. A study using IGA for the design of sounds for Electric Vehicles can be found in [14], where an exploration of the tradeoff between detectability and unpleasantness of the sounds is proposed, showing the efficiency of the approach.

2.2 Implementation of the IGA

After a definition of the variables of the sounds and their corresponding levels, a coding of the designs, represented by a chromosome, is proposed. Our implementation uses a binary coding and discrete-valued

variables. A more complete description of the implementation of our IGA can be found in [12]. The IGA creates an initial population of sounds by randomly generating the chromosomes and presents them to the user. Based on personal criteria, and according to the instructions given to the user for the experiment, the user has to rate each individual according to its "fitness". A new population of individuals is then created using one of the three operators: crossover, mutation, and selection. For each individual of the population, the random choice of the operator is controlled by the crossover rate (c_r), the mutation rate (m_r) and the selection rate (s_r). These values are chosen between 0 and 1 in such a way that $c_r + m_r + s_r = 1$. For each individual i , an indicator, $rand(i)$, is randomly chosen between 0 and 1 according to a uniform distribution:

- If $rand(i) < c_r$, the operation is a crossover (single point crossover – the second parent is randomly chosen in the population),
- If $c_r \leq rand(i) \leq c_r + m_r$, the operation is a mutation (random mutation of one variable),
- If $rand(i) > c_r + m_r$, the operation is a selection (simple duplication of the individual).

A fourth important parameter of the IGA is the roulette wheel w_r . In the crossover operation, the probability that an individual is a parent in the crossover operation is increased by the weight $w_r > 1$. An automatic process was implemented to tune the different parameters of our IGA [12]. This process uses simulated "virtual" users and a "target" product in the design space (defined by target values of the design variables). To simulate the choices of a virtual user, a distance function between the individuals of the population and the target is computed. By launching several simulations in the same conditions (Monte Carlo method), an average estimate of the convergence rates of the IGA is computed, given the value of the parameters. This process allows the experimenter to determine the "optimal" tuning of the parameters, given the maximum number of generations allowed.

3. MATERIAL AND METHODS

3.1 Synthesis of the alarm sounds

A study of different works on alarm sounds led to a list of acoustic parameters that are generally considered in perceptual studies [4-5-7]. Based on past experiments and pilot tests, 7 variables, that may have a plausible effect on the perceived severity level of an alarm, were selected. The name and definition of the levels of the different variables are given in Tab. 1. All the variables are quantitative except G (timbre) that is qualitative. G1 corresponds to a harmonic sound with 10 harmonics (decreasing magnitude in $1/n^3$), G2 corresponds to the superposition of two harmonic sounds with an

augmented fourth interval (triton), and G3 to an harmonic sound with a sawtooth waveform.

Label	Name	Levels and unit
A	Frequency of the pulse	570-900-1600 (Hz)
B	Number of bursts	2 - 3 - 5
C	Interburst duration	0.1 - 0.25 - 0.5 (s)
D	Number of pulses	2 - 3 - 4
E	Interpulse duration	0.01 - 0.1 - 0.25 (s)
F	Pulse duration	0.1 - 0.25 - 0.5 (s)
G	Timbre of the pulse	G1 - G2 - G3

Table 1. Definition of the design variables of the sounds

A 0.01s transient (onset and offset) was added to all the sounds. The representation of a typical alarm sound in the time-domain is given in Fig. 1.

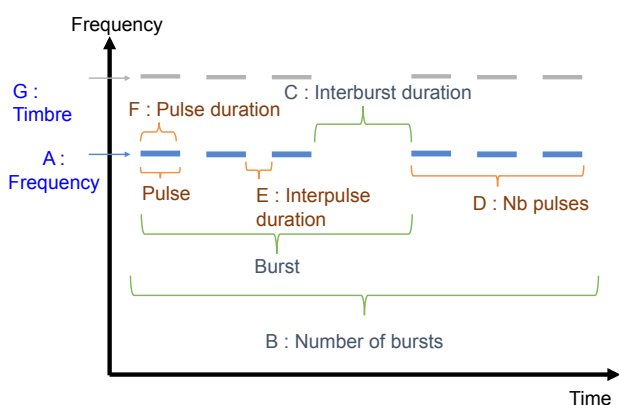


Figure 1. Illustration of the variables of the alarm sound in the time domain.

3.2 Hearing test

The objective of the hearing test was to assess the perceived severity level of alarm sounds. Participants were informed that the sounds are dedicated to assist the driving of a vehicle (e.g. warning of a low external temperature, or exceeding the speed limit) but no particular application was described.

30 participants (20 males), with no reported auditory deficiencies, performed the tests. The audio stimuli were presented with the same hardware desktop configuration, sound card and software, as well as *Beyerdynamics DT-990* headphones in a quiet environment. In the beginning of the experiment, a short tutorial was proposed, during which basic explanations on the severity level and on the use of the interface were given by the experimenter. The participants were asked to adjust the audio volume to a comfortable level and to not change this level during the experiment. After the hearing of the alarm sound, participants were asked to rate the perceived severity level on a continuous structured scale, from “not severe” to “extremely severe” (example in Fig. 2).



Figure 2. Structured scale for the assessment of the severity level.

The hearing test was divided in two sessions: **Experiment 1** (on the DOEs and with the IGA) and **Experiment 2** (hearing of the optimal sounds) few days later.

3.3 Experiment 1

This experiment consists of two parts:

- Hearing test according to a fixed set of alarm sounds, defined by an experimental design (DOE)
- Hearing test according to the IGA experiment (IGA)

All the participants did both parts (within-subject experiment) in a balanced order. The duration of the experiment was around 30mn.

3.3.1 DOE experiment

A set of alarm sounds, defined by a DOE-software, was proposed to the participants. Two models were considered to define the experimental designs: a quadratic model with interaction terms (Eqn. (1)),

$$y = \mu + a_1.A + b_1.B + c_1.C + d_1.D + e_1.E + f_1.F + \alpha_i + a_{21}.A^2 + b_{21}.B^2 + c_{21}.C^2 + d_{21}.D^2 + e_{21}.E^2 + f_{21}.F^2 + a_{17}.A.G + b_{23}.B.C + b_{24}.B.D + c_{35}.C.E + d_{45}.D.E + d_{46}.D.F + e_{56}.E.F + \varepsilon \quad (1)$$

and a linear model without interaction (Eqn. (2)),

$$y = \mu + a_1.A + b_1.B + c_1.C + d_1.D + e_1.E + f_1.F + \alpha_i + \varepsilon \quad (2)$$

with α_i : coefficient of level G_i of the qualitative variable G , with $\sum_{i=1}^3 \alpha_i = 0$. The terms considered in Eqn. (1) (number of parameters $p = 20$) were introduced according to past studies on the effect of acoustic parameters on the perception of alarm sounds.

The second model (Eqn. (2) – $p = 9$ parameters) was considered to define to which extent a simple linear model is able to represent the perceptions of the severity of alarms. With the model Eqn. (1), a DOE with 35 designs was defined (DOE-Q) according to the

optimization of the G-efficiency criterion [15] (G-efficiency = 57%). For the second experimental design corresponding to Eqn. (2), in order to limit the number of sounds to assess (limitation of the fatigue of the participant), we decided to include the 16 designs of DOE-Q located at the frontier of the experimental domain, and to complete it with a reasonable number of additional designs. A DOE with 24 designs (DOE-L) was defined according to the optimization of the G-efficiency criterion (G-efficiency = 84%). To test the lack of fit of the models, 2 repetitions of a particular design of DOE-Q (nearly a center point) were added to these DOEs.

In total, the DOE experiment counts $35+(24-16)+2 = 45$ sounds to assess, presented in a random order to the participants.

3.3.2 IGA experiment

The assessment of the alarm sound according to the severity level has been included in an IGA iterative experiment. The fitness function of the IGA was simply the severity score given by the participant (the higher the severity score, the higher the fitness). With this definition, the IGA test may converge toward sounds that represent severe events for the user. The problem of IGA tests is to manage a good balance between convergence properties and fatigue of the participant. After different tests, we considered that the assessment of 6 populations of 10 sounds was enough to not fatigue the participant (60 sounds to assess). This corresponds to a test duration of around 15mn. An automatic process was implemented to tune the different parameters of our IGA [12]. The optimal tuning parameters of the IGA are as follows:

- Wheelrate: $w_r = 25$
- Crossrate: $c_r = 0.85$
- Mutation rate: $m_r = 0.1$

The initial population of sounds was randomly generated, different for each participant.

3.4 Experiment 2

The objective of this second experiment is to compare the efficiency of “optimal” sounds (optimal according to severity) defined from data of experiment 1. The choice of this objective function is of course arbitrary: we should have chosen to optimize according to the least severe sound, or even to take into account multiple objectives, for example severity and unpleasantness (as in [14]). This choice is not important given that we focus mainly on the methodological aspects.

One week after experiment 1, after the processing of the data of the two tests, the same participants were asked to rate a set of “optimal” sounds according to severity (similarly to experiment 1 (Fig. 2)), defined

from their ratings. In this paper, we limit the presentation to the following six sounds¹:

- The individual sound (participant i) with the highest severity score, given by the DOE-Q model: $DOE_Q_i^{opt}$ (quadratic model),
- The individual sound with the highest severity score, given by the DOE-L model: $DOE_L_i^{opt}$ (linear model),
- The individual sound with the highest severity score, given by the IGA test: IGA_i^{opt} ,
- The global sound with the highest severity score, given by the DOE-Q model fitted to the average participant: $DOE_Q_g^{opt}$ (quadratic model),
- The global sound with the highest severity score, given by the DOE-L model, fitted to the average participant: $DOE_L_g^{opt}$ (linear model),
- The global sound with the highest severity score, given by the IGA test and a selection process (see below): IGA_g^{opt} .

The fitting of the models to the “average participant” was made after a verification of a large enough consensus between the participants (consonance analysis, see below).

To define the optimal global sound for the IGA test, several procedures were considered. We choose to report in this paper only the results of the following process: given the individual optimal sound of each participant IGA_i^{opt} , the occurrences of each levels of each variable were computed. The global optimal sound IGA_g^{opt} was defined as the design with the largest occurrences for each variable (most occurring level).

In addition to the six sounds defined, two sounds, considered as “not severe” by the experimenter (LOW1 and LOW2), were added to the set of sounds.

3.5 Analysis of the data

3.5.1 Inter-subjects agreement (consonance analysis)

Before fitting a model to the “average participant”, it is necessary to verify that the panel is consensual enough (otherwise, poor results, not representative of the assessments, may be obtained). The verification of the agreement between participants is made by consonance analysis, a method based on a principal component analysis (PCA) of the assessments [16]. Observations of this PCA correspond to the sounds of the DOE, whereas variables correspond to the subjects. The principle of the method is to assess the deviation from unidimensionality

¹ Additional sounds were proposed, not reported in this paper

by an examination of the variance accounted for by each component of the PCA. A consonant panel should obtain high variance on the first component, and low variance on the other components (under the condition that the variable points are on the same side of the first component). Two standardized PCA are carried out on each DOE (DOE_Q and DOE-L) and the percentage of variance on the two first components is examined.

3.5.2 Statistical tests on the ratings of experiment 2

Differences in the ratings of the set of sounds (experiment 2) are analyzed with a linear mixed model (equivalent in our case to an ANOVA) [17]. Two factors are considered in the model, a factor “sound” with a fixed effect and factor “subject” with a random effect. Paired comparison tests between sounds are next performed using a Duncan multiple comparison tests (post-hoc).

3.5.3 Agreement IGA/DOE

For each participant, his/her DOE model can be used to compute the predicted severity scores on the entire design space (full factorial design). In particular, the severity score of the sound selected in the IGA test, IGA_i^{opt} , can be computed with the DOE model of participant i . A high severity score of the IGA_i^{opt} product is a sign of an agreement between the two tests, whereas a low severity score represents a disagreement. It has to be mentioned that this agreement between the tests includes the effect of the method and also the consistency of the participant between the two tests: if the participant radically changes his/her opinion about the severity between the two tests, the agreement may be low. To measure the assessments’ consistency between the two tests, the proposed method consists in positioning the predicted score of the IGA_i^{opt} product in the total distribution of the scores of the complete design space (we choose the full factorial of $3^7=2187$ designs). For each participant, the predicted severity scores \hat{Z}_{full}^i of the 2187 designs were computed using the 30 individual DOE models and the global models. After a ranking of the designs in an increasing order according to the score \hat{Z}_{full}^i , the IGA final choice IGA_i^{opt} is located in this ranking. The individual discrepancy between the two tests for participant i is characterized by the agreement $Ag^i \%$ (Eqn. (3)), corresponding to the percentage of designs with a lower score than the IGA_i^{opt} sound:

$$Ag^i \% = 100 \cdot \frac{Rank(\hat{Z}_{full}^i IGA_i^{opt})}{2187} \quad (3)$$

4. RESULTS

4.1 Inter-subjects agreement (consonance analysis)

The results of the standardized PCA on the subjects’ severity ratings (subject S1 to S30) are presented in Fig. 3 (DOE for the quadratic model DOE-Q) and in Fig. 4 (DOE for the linear model DOE-L).

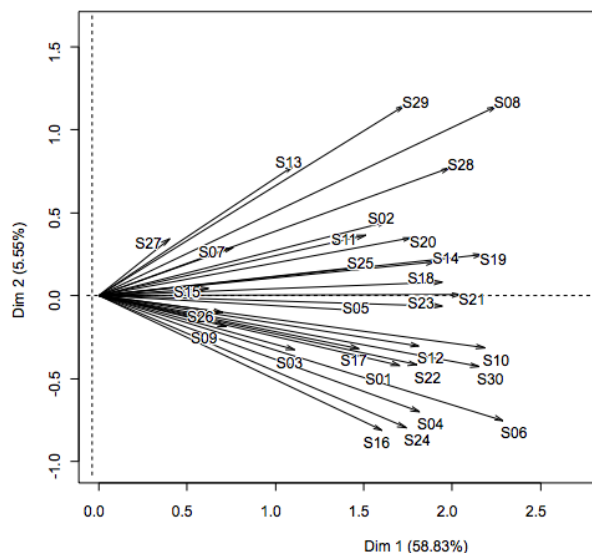


Figure 3. PCA of the ratings for the quadratic model (DOE-Q) (plane of the variables)

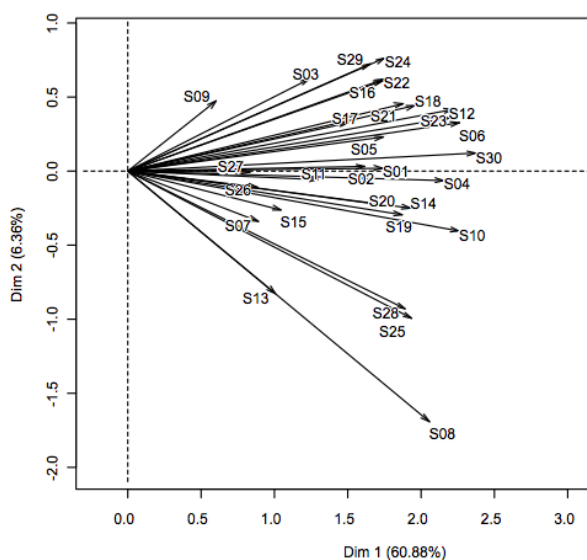


Figure 4. PCA of the ratings for the linear model (DOE-L) (plane of the variables)

The percentage of variance on the first component is very high (58.8% for DOE-Q and 60.8% for DOE-L), whereas the percentage on the second component is very low (5.5% for DOE-Q and 6.3% for DOE-L). All the arrows point in the same direction, the panel can be considered as consensual enough and a global model can be fitted to the “average participant”.

4.2 Analysis of the DOE models

4.2.1 Fitting of the models

The determination coefficients R^2 of the individual DOE models are generally large, above 70%, except for three subject S15, S27 and S29 that got R^2 between 50 and 60%. The lack of fit test is generally not significant. The determination coefficients R^2 are large for the global models DOE-Q (97%) and DOE-L (92%).

4.2.2 Effect of the variables on the severity level

The coefficients of the global DOE models that are significant ($p < .05$) are given in Tab. 2.

	DOE-L (linear)	DOE-Q (quadra)
R^2	92%	97%
μ	4.50	5.27
A	0.37	0.41
B	0.48	0.50
C	-0.22	-0.20
D	0.54	0.49
E	-0.42	-0.43
F	1.30	1.25
G1	-0.64	-0.71
G2	0.24	0.24
G3	0.40	0.47
C*E	-	0.21
A^2	-	-0.53

Table 2. Significant coefficients ($p < .05$) of the global DOE models.

There is a large agreement between the two models on the effect of the acoustic variables on the severity score:

- Increasing the frequency A, the number of bursts B, the number of pulses D, the pulse duration F increases the severity,
- Decreasing the interburst duration C, the interpulse duration E, increases the severity,
- The timbre G3 (harmonic sound with a saw tooth waveform) corresponds to the largest severity.

These results are rather in agreement with previous studies on the perceived urgency of alarm sounds [4-5-7], even if the definition of the sounds and the perceived dimension studied in our study are slightly different. This agreement strengthens the validity of our experimental protocol and the ability of our panel to produce consistent ratings.

Concerning the quadratic model DOE-Q, only two additional terms, the interaction C*E and the term A^2 , are significant. Their effect is very slight and does not modify the previous conclusions. We have to conclude that the linear model is sufficient to represent the data, and that the quadratic model is in overall not necessary.

4.3 Agreement between the IGA and DOE tests

For each participant and for the global models, the percentages of agreement $Ag\%$ between the IGA and the DOE tests are given in Fig. 5, for both models DOE-L and DOE-Q.

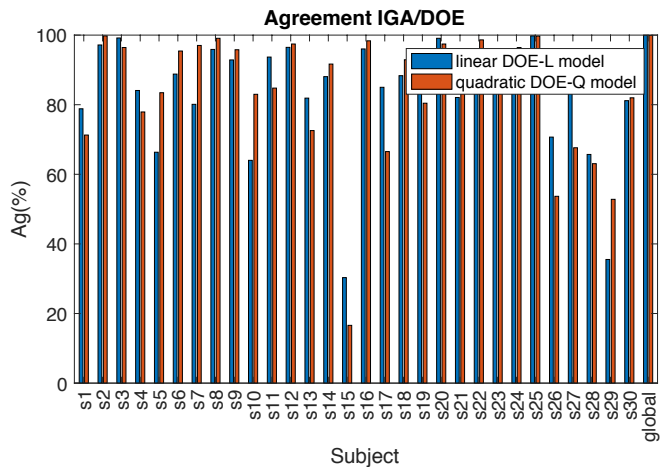


Figure 5. Percentage of agreement $Ag\%$ between the IGA and the DOE tests for all the subjects (S1 to S30) and the global model.

The agreement is important, above 60% for a majority of participants. The average agreement (across participants) is 83% for the DOE-Q and DOE-L models. The quadratic model does not improve the agreement, sign that confirms that the linear model is satisfactory to represent the data. Two subjects S15 and S29 obtain low $Ag\%$ (lower than 50%). It is interesting to mention that the fitting of the DOE models (R^2) of these subjects was also very low (around 50%), making their assessments suspicious if not inconsistent.

Overall, the individual results of the participants are rather consistent between the two tests. The confidence in the data is large, except for the two subjects S15 and S29. For the global model, the $Ag\%$ is 100% for the DOE-L model ($DOE_L_g^{opt}$ sound is exactly the same than IGA_g^{opt}) and 99.95% for the DOE-Q model ($DOE_Q_g^{opt}$ sound is slightly different (difference of one level on only two variables) of IGA_g^{opt}). At the group level, the two tests lead to similar conclusions concerning the most severe sounds, the agreement between the results of the two tests is very important.

These results show that the two approaches (DOE and IGA) lead to similar conclusions concerning the most severe sound. The IGA method converges and the experimental protocol designed is satisfactory.

4.4 Comparison of the optimal sounds

Given that the two optimal sounds IGA_g^{opt} and $DOE_L_g^{opt}$ are exactly the same, only 5 optimal sounds are considered for comparison.

The ratings of these 5 sounds plus the sounds LOW1 and LOW2 by the 30 participants according to the severity score (experiment 2) were analyzed with a linear mixed model. The results of the analysis give a

significant effect for the sounds ($F(6,174)=89.5 - p<.0001$) and a significant effect for the participant ($(F(29,174)=6.5 - p<.0001)$). The results of the post-hoc tests with the Duncan multiple comparisons test are presented in Fig. 6.

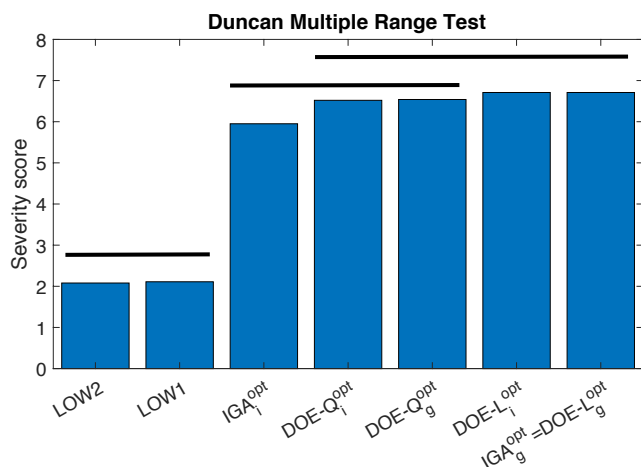


Figure 6. Duncan multiple comparisons test. Non-significant pairs ($p>.05$) are connected with a continuous line.

The results show first that the two sounds LOW 1 and LOW 2, considered as particularly not severe by the experimenter, are also perceived as such by the panel of participants. This confirms the validity of the participants' ratings. The differences between the optimal sounds are very thin, the only slight significant difference concerns the sound IGA_1^{opt} that is perceived as less severe than two other sounds (a PCA (not reported here) showed that this fact is due to a subgroup of 3 subjects (S15 S16 S19) that assessed their optimal sound IGA_1^{opt} as less severe than the other optimal sounds. A first conclusion is that the IGA method does not allow the participant to design a sound that is significantly most severe than the other optimal sounds. At most (if subjects S15 S16 S19 are discarded), sounds are perceived as equivalent. This result, rather disappointing, could be explained by the weak number of generation allowed for the IGA method (6 generations only) and also could be due to the convexity of the objective function (if the optimum is unique, all methods converge to it).

A second conclusion is that at the group level, the different methods and models considered in the study provide equivalent results. No method outperforms the other in the definition of an optimum.

For information, the definition of the two optimal global sounds ($IGA_g^{opt} = DOE-L_g^{opt}$) and $DOE-Q_g^{opt}$ are given in Tab. 3. These two sounds are very similar, the only differences concern the frequency (1300Hz instead of 1600Hz for the quadratic model) and the number of bursts (4 instead of 5 for the quadratic model).

Var.	Name	$IGA_g^{opt} - DOE-L_g^{opt}$	$DOE-Q_g^{opt}$
A	Frequency of the pulse	1600 Hz	1300 Hz
B	Number of bursts	5	4
C	Interburst duration	0.1s	0.1s
D	Number of pulses	4	4
E	Interpulse duration	0.01s	0.01s
F	Pulse duration	0.5 s	0.5 s
G	Timbre of the pulse	G3	G3

Table 3. Definition of the optimal sounds (largest severity) at the group level

To be perceived as severe, an alarm sound must, with the definition of our variables, have a high frequency (A), a large number of bursts (B) and of pulses (D), a short interburst (C) and interpulse (E) duration, a large pulse duration (F) and the timbre G3 (sawtooth waveform). This is in agreement with previous studies.

4.5 Discussion

In this paper, our objective was to study the ability of IGA to design optimal sounds (with the largest severity) in a perceptual study using hearing tests. The results show that at the individual level, the IGA method does not allow the participant to design a sound that is very different to other proposal obtained by the classical DOE method. This result a little disappointing is perhaps due to the choice of the sound parameters and the perceived dimension under study (the severity of alarms) that make the objective function convex. Further studies should be designed to investigate this aspect.

The experiment also does not allow the highlighting of the superiority of a model-free method in the integration of interactions effect in the definition of optima. Our objective in the beginning of the experiment was to fit two models, a linear model (DOE-L) and a quadratic model (DOE-Q) in order to show that the optima obtained by the IGA method would be closer to the optima of the quadratic model than the optima of the linear model. We were not able to prove this given that no interaction nor square effect were significant and very important in our perceptual study. Further studies where clear interactions between variables occur should be designed to prove this advantage.

Nevertheless, this study provides interesting outcomes: the IGA method is at least as efficient as the DOE approach, for a number of ratings that is comparable (45 sounds for the DOE-Q model, 60 sounds (6 generations of 10 sounds) for the IGA). We could use the DOE approach to rough out the main variables of a design problem, important for a perceptual study, and

use in a second stage the IGA method to optimize the design on a low number of variables, where interaction and quadratic effect could be influential.

5. CONCLUSIONS

We showed in this paper that hearing tests based on IGA are an interesting method to optimize a given perceptual dimension (the severity level of alarm sounds) and to define relevant sounds. It can be a valuable alternative to classical design of experiments (DOE), in particular when no model is a priori available between the design variables and the perceptual dimension. On the case study designed (severity of alarm sounds), we showed that the quality of the results given by the IGA method is equivalent to those of the approach using DOE with a quadratic or a linear model.

Further studies are needed to prove the superiority of IGA when important interactions between the design variables occur. In addition, this method can be used to customize sounds and adapt to the customer taste. This will be the purpose of next experiments.

6. REFERENCES

- [1] N.A. Stenton and J. Edworthy. Auditory warnings and displays: an overview. In N.A. Stenton & J. Edworthy (Eds). *Human factors in auditory warnings* (pp. 3-30). Aldershot, UK: Ashgate, 1999.
- [2] V. Roussarie, F. Richard, M.C. Bezat "Perceptive qualification of engine sound character; validation of auditory attributes using analysis-synthesis method", *Proc. of CFA/DAGA'2004*.
- [3] J. Edworthy, S. Loxley and I. Dennis. "Improving auditory warning design: relationship between warning sound parameters and perceived urgency. *Human Factors*, 33, 205 – 231, 1991.
- [4] D. C. Marshall, J. D. Lee and P. A. Austria. "Alerts for In-Vehicle Information Systems: Annoyance, Urgency, and Appropriateness". *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(1), 145–157, 2007.
- [5] C. Suied, P. Susini and S. McAdams. "Evaluating Warning Sound Urgency With Reaction Times". *Journal of Experimental Psychology: Applied*, 14(3), 201–212, 2008.
- [6] B. Sousa, A. Donati, Elif Ozcan Vieira, Rene van Egmond, Reinier Jansen, J Edworthy, R. Pelszus, Y. Voumard. Designing and deploying meaningful auditory alarms for control systems. *Proc. of SpaceOps 2016*, Korea, 2016.
- [7] J. P. Singer, N. D. Lerner, C. L. Baldwin and E.C. Traube. "Auditory Alerts in Vehicles: Effects of Alert Characteristics and Ambient Noise Conditions on Perceived Meaning and Detectability". *Engineering*, 2015.
- [8] Takagi, H. 2001. Interactive Evolutionary Computation: Fusion of the Capabilities of EC Optimization and Human Evaluation. *Proc. IEEE*, 89, 9, 1275–1296.
- [9] M. Miki, H. Orita, S. H. Wake and T. Hiroyasu, "Design of Sign Sounds using an Interactive Genetic Algorithm," *2006 IEEE International Conference on Systems, Man and Cybernetics*, Taipei, 2006, pp. 3486-3490, doi: 10.1109/ICSMC.2006.384659.
- [10] D. E. Goldberg. *Genetic Algorithms in search, optimization & machine learning*, Addison-Wesley Publishing Company, 1989.
- [11] J. H. Lee and M. L. Chang. "Stimulating designers' creativity based on a creative evolutionary system and collective intelligence in product design". *International Journal of Industrial Ergonomics*, 40, 295–305, 2010.
- [12] E. Poirson, J-F. Petiot, L. Boivin and D. Blumenthal. "Eliciting User Perceptions Using Assessment tests based on an Interactive Genetic Algorithm". *Journal of Mechanical Design*, Vol. 135, Issue3, 031004 -1, 2013.
- [13] G.H. Wakefield, C. Van den Honert, W. Parkinson and S. Lineaweaver. "Genetic Algorithms for Adaptive Psychophysical Procedures: Recipient-Directed Design of Speech-Processor MAPs". *Ear and Hearing*, 26(4) - p 57S-72S, 2005.
- [14] J-F Petiot, K. Legeay and M. Lagrange. "Optimization of the sound of electric vehicles according to unpleasantness and detectability". *Proc. of ICED 2019*, Delft, 2019.
- [15] W. Kuhfeld. "Efficient Experimental Designs Using Computerized Searches", *Sawtooth Software Conference Proc.*, 1997.
- [16] G. Dijksterhuis. Assessing Panel Consonance, *Food Quality and Preference*, vol. 6, pp. 7-14, 1995.
- [17] A.I. Khuri, T. Mathew, B.K. Sinha. *Statistical Tests for Mixed Linear Models*, John Wiley, New York, 1998.