



HAL
open science

Collaborative Robotics: Application of Delphi Method

Jean-marc Salotti, Ephraim Suhir

► **To cite this version:**

Jean-marc Salotti, Ephraim Suhir. Collaborative Robotics: Application of Delphi Method. Journal of Field Robotics, 2024, 10.1002/rob.22477 . hal-04868639

HAL Id: hal-04868639

<https://hal.science/hal-04868639v1>

Submitted on 11 Feb 2025

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.

RESEARCH ARTICLE OPEN ACCESS

Collaborative Robotics: Application of Delphi Method

Jean-Marc Salotti^{1,2}  | Ephraim Suhir^{3,4,5}

¹Auctus Team, Inria, France | ²IMS Laboratory, CNRS UMR 5218, Bordeaux INP-ENSC, Bordeaux University, Talence, France | ³Portland State University, Portland, Oregon, USA | ⁴University of California, Los Angeles, Los Angeles, California, USA | ⁵ERS Co, Los Altos, California, USA

Correspondence: Jean-Marc Salotti (jean-marc.salotti@ensc.fr)

Received: 29 June 2024 | **Revised:** 24 September 2024 | **Accepted:** 3 November 2024

Keywords: decision making | Delphi method | multicriteria analysis | teleoperated robot

ABSTRACT

In the domain of collaborative robotics, it is often needed to make complex multiparametric choices or assessments that combine heterogeneous criteria, such as task performance, reliability, safety, and ergonomics. The Delphi method, a process of arriving at group consensus by providing experts with rounds of questionnaires, as well as the group response before each subsequent round, can be applied to solve this type of problems. The following Delphi method efforts are addressed in our analysis with application to finding appropriate criteria and their relative weights for the evaluation of the teleoperation task of a robotic platform: (1) Contribution of experts; (2) estimates using student's distribution; and (3) decision making support using Fischer criterion. A consensus emerged among eight experts, who selected the criteria of duration, number of trajectory corrections, and feeling of the operator. Weights have been proposed for each criterion and statistical tools have been used to evaluate the consensus.

1 | Introduction

The design of collaborative robots is a critical and extremely complex task (Goodrich and Schultz 2007; Mutlu and Roy 2016; Sheridan 2016; Demir et al. 2019; Coronado et al. 2022; Apraiz et al. 2023). An ideal solution would be using a robot replacing the human operator. However, this is hardly practical, both because the environment and/or the task are too complex, and because there is a permanent need for a proper and timely adaptation/adjustment of the governing factors. As a consequence, it is not usually possible, nor advisable to avoid employing human operators. In such a situation, partial automation of the “human-in-the-loop” type of situation (Suhir 2018), when humans and robots interact and collaborate, looks like the inevitable and the best solution. Many natural questions arise because of that for the design of a cobotic system (Moulières-Seban et al. 2016; Villani et al. 2018). Examples are: how to split the critical task into subtasks? how to determine the best allocation for the robot and the humans? how to determine the most appropriate type of interactions between them? or what will be the, in effect, never-zero risks of possible accidents (Cambor et al. 2021)? Furthermore, once a robot is built and the operation

interface defined, how to assess the system's quality (Apraiz et al. 2023; Coronado et al. 2022)? What criteria have to be considered? and how to combine them into a global assessment function? The use of the Delphi method is suggested in this write-up to at least partially answer these questions: statistical tools are used to benefit from the knowledge and experience of experts to find appropriate answers. Both the Delphi method is presented and the associated mathematical tools. An application of a tele-operated robotic platform for the transportation of heavy loads is addressed as a suitable illustration.

2 | Delphi Method

2.1 | Background: The Delphi Method (DM)

Useful information of a very broad nature can be often obtained, as is known, in a timely and cost-effective fashion about an event, an issue, a possibility, a situation, a methodology, or a parameter of interest by processing, in a systemic way, opinions (feedback) of a selected group of knowledgeable experts, using DM (Brown 1968; Dalkey 1969; Adams 1980; Bolongaro 1994;

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Author(s). *Journal of Field Robotics* published by Wiley Periodicals LLC.

Koynak et al. 1994). In some applied science and engineering undertakings such an approach could be implemented instead of conducting costly and time-consuming actual experimentation (Brown 1972; Azani and Khorramshahgol 1990; Mitchell 1992; Breiner et al. 1994; Suhir 1996; Lazarus et al. 2022). Other methods try to determine the best consensus among experts by providing feedback. For instance, Weidong et al proposed an interaction based on trust relationship and a Cooperation Degree Index (Weidong et al. 2024). In another approach, Sun et al. suggested the use of a Trust Consensus Index to assess the discordance of experts and apply weight penalty mechanisms (Sun et al. 2023). To avoid trust issues and cognitive biases caused by pre-existing human relationships, the DM coordinator may impose an anonymous expression. For that reason, the DM is the best-known and the most widely used technique of this type. The method has been proven to be useful in numerous decision-making problems in various fields of forecasting, auditing, planning, and business management (Brockhaus and Mickelson 1976; Goldfisher 1992; Suganthi and Jagadeesan 1992; Linstone and Turoff 2002). The participating experts do not actually interact: they express their opinions independently, by answering sets of highly focused questions from the previously developed questionnaires. The information obtained from a particular expert is treated as an independent sample of a random (uncertain) variable (characteristic) of interest, and more or less sophisticated statistical methods for analyzing these random data are being subsequently used to process and interpret the obtained information. The DM can be used as a probabilistic predictive modeling technique that assumes that the deviations of expert opinions (estimates) from the “objective truth” are due to random causes, and that the application of the DM is able to “restore” such an “objective truth” with a certain and significant degree of confidence.

The DM assesses the meaningfulness of the estimates, the degree of agreement of the experts' opinions, and can be even used to select and form the right team of experts by assessing their suitability and levels of expertise. After the experts' estimates are processed, each expert might be given the resultant estimate. The experts might be provided also with some additional information. Using these data, he/she might revisit and re-examine his/her estimates. The entire procedure is repeated until the experts' estimates become reasonably consistent. The method is most suitable, when there is a need to provide an insight into an undeveloped subject area, when a long-range forecast is needed, when the past data are not relevant to the future, and when consistent historical data do not exist or are not available. The major strengths of the DM are the use of sound reasoning; applicability to a wide range of issues; absence of interference from group social pressures; and ease in bringing in experts. The DM's drawbacks are, in effect, common to all judgmental forecasts: judgmental nature of such forecasts; difficulties in selecting suitable experts; evaluating their expertise and other human qualities important for obtaining an objective and consistent information; difficulties in estimating how accurate the obtained information is; numerous challenges in creating effective and goal-oriented questionnaires. The typical problem and complaint is the difficulty of evaluating the actual expertise and/or the conscientiousness of the experts. In the analysis below, we show how three more or less well-known methods of processing DM data can be applied in some

problems in the domain of human-robot collaboration. The following particular problems are addressed: (1) Selection of experts (Section 2.2); (2) DM-based estimates using Student's distribution (Section 2.3); (3) DM-based decision using Fisher criterion (Section 2.4).

2.2 | Selection of Experts

Before starting questioning experts using the developed questionnaire, the DM analyst has to determine the actual expertise of the experts, the thoroughness of their thinking, their “human capacity factors” (HCF) (DeHaan and Peters 1993; Helmer 1994; Backley 1995; Suhir 2011, 2018), and the appropriate number of experts. Apart from competence in the field, a suitable expert should possess some important additional qualities: creativity, that is, the ability to address problems, for which methods of solution are unknown; heuristic ability, that is, the ability to identify nontrivial problems; good intuition (“gut feeling,” “educated guess”), that is, the ability to see a solution without knowing why; anticipation, that is, the ability to foresee a solution or a potential pitfall; independent mind, that is, the ability to withstand the opinions of others, even if these opinions are in the overwhelming majority; versatility, that is, the ability to see a problem from a different point of view than suggested by the DM director. Formalized submitted information about the source for the experts' answers and the degree of influence due to each source on his/her answer can be obtained, for example, by asking the experts to complete a table of the Table 1 type below.

The number of experts should be sufficient to ascertain that the essential features of the problem are captured. Too many experts may lead to significant differences of opinion, as well as to difficulties in the organization of the expert procedure and processing of the obtained data (“too many cooks spoil the broth”). It is advisable that the group of experts consists of not less than 20 and not more than 50 individuals. Once the number of experts is decided upon, the selection of particular individuals can be carried out in several stages. For instance, a small group of specialists capable of providing opinions could be initially nominated by the project manager (coordinator). Then the questionnaires are given to each of the specialists named. They, in turn, give the names of specialists competent in the topics under consideration. The process continues until new names do not appear anymore. Such a procedure will most likely result in a complete list of the existing, known, trustworthy, and competent persons in the organization (agency/field). The questions in a questionnaire should be formulated so as to exclude ambiguous answers. The questionnaire should

TABLE 1 | Analysis of expertise.

Source of expert's answer	Influence of an expert		
	High	Average	Low
Own analysis	X		
Own experience	X		
Studies of others		X	
Intuition			X

involve questions requiring qualitative answers (“better/worse,” “more/less,” “higher/lower”), even if the ultimate goal of the DM undertaking is to obtain quantitative information.

Example: Do you think that an accident with a given robot is impossible, highly unlikely, unlikely, less likely than more, has a 50%/50% chance, more likely than less, likely, highly likely, almost certain (circle the appropriate answer). These responses could be interpreted as 0%, 12.5%, 25.0%, 37.5%, 50.0%, 62.5%, 75.0%, 87.5%, and 100% probability that the accident will occur. The questions in the questionnaires should be arranged hierarchically, that is, more general questions should be asked first, and more specific questions should follow.

2.3 | Estimate Using Student's Distribution

Student's t -distribution is a family of continuous distributions that arises when estimating the mean values of a normally distributed random variable in situations where the sample size is small and the standard deviation of the population is unknown (Suhir 1997). Student's t -distribution is symmetric and bell-shaped, like the normal distribution, but has “heavier tails,” meaning that it is more prone to producing values that fall far from its mean. This makes it useful for understanding the statistical behavior of certain types of ratios of random quantities, in which variation in the denominator is amplified and may produce outlying values when the denominator of the ratio falls close to zero. In the example that follows we assess, using DM data and Student's t -distribution the most likely value of the duration of the anticipation time. The general procedure is as follows (Suhir 1996, 1997).

Let a parameter X , such as, for example, in the case in question, the weight of a criterion in the formula used to evaluate the teleoperation task of a robotic platform, be estimated by N independent experts. The effective estimate α of the parameter X is:

$$\alpha = \frac{\sum_{k=1}^N \alpha_k x_k}{\sum_{k=1}^N \alpha_k}. \quad (1)$$

Here α_k is the competence (“weight”) of the k -th expert, and x_k is the estimated X value by this expert. The extent of disagreement among the expert opinions about parameter X can be characterized by the normalized variance:

$$D = \frac{\sum_{k=1}^N \alpha_k (x_k - \alpha)^2}{\sum_{k=1}^N \alpha_k}. \quad (2)$$

The statistical significance of the obtained result can be defined as

TABLE 2 | Example.

k	1	2	3	4	5	6	7	8	9	10
x_k	33.0	35.0	32.2	34.0	38.0	34.0	37.0	40.0	36.0	35.5

$$\bar{\alpha} - \Delta \leq \alpha \leq \bar{\alpha} + \Delta, \quad (3)$$

in which the value of interest can be found with the given (required, desired, specified) probability of $1 - Q$, where Q is the probability of error. In the above inequality (3), $\bar{\alpha}$ is the mean value of the estimate α and Δ is the deviation of the actual (random) value of the estimate α from its mean value. If the variable α is normally distributed with the mean value $\bar{\alpha}$ and the variance D then the deviation of the actual value of the estimate α from its mean value is

$$\Delta = \frac{t\sigma}{\sqrt{N}}, \quad (4)$$

where the t value has Student's distribution with $N - 1$ degrees of freedom, and $\sigma = \sqrt{D}$. Let the experts' estimates of the number of seconds that a robotic operator needs to anticipate a particular situation of importance during a certain teleoperation task are as shown in Table 2:

All the experts have the same “weight” $\alpha_k = 1$, the acceptable probability of error is $Q = 5\%$, and one wants to establish the confidence interval for the estimate $\alpha = 35.5$ determined by Equation (1) and the degree of disagreement $D = 4.90$ found from Equation (2). The tables for the Student's distribution for the number $N - 1 = 9$ of degrees of freedom, the probability of error $Q = 0.05$, and the standard deviation $\sigma = \sqrt{D} = \sqrt{4.90} = 2.2136$, yield: $t = 2.262$. Then Equation (4) yields: $\Delta = 1.583$. From the inequality (3), we conclude that the actual value of the duration of the anticipation time is between 33.917 and 37.083 s, with the probability of 95%.

2.4 | Decision Making Support Using Fisher Criterion

The DM can also be applied for the evaluation, using opinions of experts (reviewers), of a particular choice (proposal, strategy, approach, technology, plan, item, product, mission, initiative, service, software, risk, interface, etc.), and for making a decision (choice) based on such an evaluation (Suhir 1996). First, the experts are requested to fill out questionnaires aimed at the selection of the choice to be made. Then they are requested to answer another questionnaire, in which, based on the results of the analysis of the first one, the selected characteristics are ranked, that is, are listed in order of their significance. The objective of the third questionnaire is to apply the established characteristics to evaluate the possible choices and to make the decision about the best choice that possesses the most valuable characteristics. Here is how it could be done.

Let the number of experts be n ($i = 1, 2, \dots, n$), the number of characteristics of the choice under consideration be m ($j = 1, 2, \dots, m$), and the significance numbers in the developed

ranking matrix be x_{ij} , where i refers to an expert and j to a particular characteristic. The sums of all the columns and all the lines in the ranking matrix should satisfy the condition:

$$\sum_{j=1}^m \sum_{i=1}^n x_{ij} = \sum_{i=1}^n \sum_{j=1}^m x_{ij}. \quad (5)$$

Using the ranking matrix, one can assess the degree of correlation of the expert opinions. The discrepancies can be due to different qualifications of the experts and/or to their different opinions about the particular characteristic because of different knowledge about this characteristic. If there are no equal elements (ranks) x_{ij} in the ranking matrix, the concordance coefficient C (also known as Kendall's W) can be calculated as

$$C = \frac{12S}{n^2m(m^2 - 1)}, \quad (6)$$

where

$$S = \sum_{j=1}^m \left(\sum_{i=1}^n x_{ij} - \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n x_{ij} \right), \quad (7)$$

is the correlation factor (Suhir 1997). If some of the elements (ranks) in the ranking matrix are the same, then the formula

$$C = \frac{12S}{n^2m(m^2 - 1) - n \sum_{i=1}^n T_i}, \quad (8)$$

for the concordance coefficient should be used. In this formula

$$T_i = \sum_{j=1}^m t_j \left(t_j^2 - 1 \right), \quad (9)$$

and t_j is the number of ranks of the j -th type in each of the lines of the ranking matrix. If the calculated C value is small (say, smaller than 0.1) this could be an indication that the experts have been chosen wrongly and/or that their knowledge about the subject is not sufficient. If the calculated C value is very large (say, larger than 0.9), one should conclude that the analysis has been carried out too superficially, too formally, without an in-depth study of the attributes of the object of interest. In both extreme cases, the process should be revisited and repeated.

The deviation of the concordance coefficient from zero can be checked using the F test or Fisher criterion:

$$F = \frac{W(m - 1)}{1 - W}. \quad (10)$$

The calculated F value is compared with the F_α value determined for a low level $\alpha = 0.01 - 0.05$, and the degrees of freedom ν_1 and ν_2 . These are calculated as $\nu_1 = m - 1 - \frac{2}{n}$, $\nu_2 = (n - 1)\nu_1$. If $Z < Z_\alpha$, then one should conclude, with the

probability $Q \geq 1 - \alpha$, that there is no agreement among the experts. In this case, one should conduct new analysis or substitute the experts with new ones, who would agree better. This can be done, for example, by excluding one of the experts from the team and by evaluating the coefficient C_1 for the remaining experts. If $C_1 > C$, this expert should be excluded from the team, and if $C_1 < C$, then he/she should remain on the team. Such evaluations should be conducted for each expert. As a result of these evaluations, the degree of agreement among the experts remaining on the team increases.

At the next step, one assesses the difference in the significance of different characteristics. This can be done on the basis of the computed variances. The significance of the rank x_{ij} is due to the three independent components (inputs): (1) the component (input) of the given expert; (2) the component (input) that is due to the given characteristic; and (3) the remainder that can be treated as a normally distributed random variable with zero mean and nonzero/finite variance.

Hence, the total variance of the given characteristic can be represented as a sum of three components:

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x}_{ij})^2 &= \sum_{i=1}^n \sum_{j=1}^m (\bar{x}_i - \bar{x}_{ij})^2 + \sum_{i=1}^n \sum_{j=1}^m (\bar{x}_j - \bar{x}_{ij})^2 \\ &+ \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x}_{ij})^2. \end{aligned} \quad (11)$$

Here,

$$\bar{x}_{ij} = \frac{1}{mn} \sum_{i=1}^n \sum_{j=1}^m x_{ij} \quad (12)$$

is the total mean rank,

$$\bar{x}_j = \frac{1}{m} \sum_{j=1}^m x_{ij} \quad (13)$$

is the mean rank for the i -th expert, and

$$\bar{x}_i = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (14)$$

is the mean rank for the i -th expert. The assessment of the differences in the significances of different characteristics should be carried out by comparing the variances

$$D_1 = \frac{1}{m - 1} \sum_{i=1}^n \sum_{j=1}^m (\bar{x}_i - \bar{x}_{ij})^2 = \frac{1}{m - 1} \left(n \sum_{j=1}^m \bar{x}_i^2 - mn \bar{x}_{ij}^2 \right) \quad (15)$$

between the different characteristics, with the remaining variance

$$\begin{aligned} D_r &= \frac{1}{(m - 1)(n - 1)} \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x}_{ij} - \bar{x}_j + \bar{x}_{ij})^2 \\ &= \frac{1}{(m - 1)(n - 1)} \left(\sum_{i=1}^n \sum_{j=1}^m x_{ij}^2 - m \sum_{i=1}^n \bar{x}_i^2 - n \sum_{j=1}^m \bar{x}_j^2 + mn \bar{x}_{ij}^2 \right). \end{aligned} \quad (16)$$

The significance of the difference in the variances D_1 and D_r can be evaluated on the basis of the following Fisher criterion:

$$Z = \frac{1}{2} \ln \left(\frac{D_1}{D_r} \right), \quad (17)$$

for the degrees of freedom ν_1 and ν_2 , calculated as $\nu_1 = m - 1$, and $\nu_2 = (m - 1)(n - 1)$. If $Z \geq Z_\alpha$, where Z_α is determined for a low enough α value (say, $\alpha = 0.05$), then one concludes that the difference between the variances D_1 and D_r is significant. This means that the difference in the roles of different characteristics is substantial, and the influence of the chosen characteristics is significant. If, however, $Z < Z_\alpha$, then the difference between the variances D_1 and D_r is insignificant, the difference in the roles of different characteristics is small and the influence of the given characteristic is not essential. If this is the case, one should broaden the number of characteristics and start a new Delphi process.

When conducting statistical analyses, one might be willing to assess the probability distributions of different characteristics and the differences in these distributions. The assessment of the significance in the distributions of different characteristics can be substituted by the comparison of the mean values of the corresponding ranks. Let us examine now how the difference in the mean values of the ranks can be assessed.

When the ranks for each criterion from random samples are evaluated, each of these samples is characterized by its mean \bar{x}_{ij} value and variance,

$$D_j = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_{ij})^2. \quad (18)$$

The significance of the difference in the mean values of different ranks for different characteristics can be evaluated on the basis of the Student distribution tables by using a sequential comparison of the mean ranks for different characteristics:

$$t_{\bar{x}_k - \bar{x}_l} = \frac{\bar{x}_k - \bar{x}_l}{\sqrt{D_{\bar{x}_k - \bar{x}_l}}}, \quad k, l = 1, 2, \dots, m, \quad k \neq l, \quad (19)$$

and the variance $D_{\bar{x}_k - \bar{x}_l}$ can be computed as

$$D_{\bar{x}_k - \bar{x}_l} = D_{\bar{x}_k} + D_{\bar{x}_l}, \quad (20)$$

and

$$D_{\bar{x}_j} = \frac{D_j}{n}, \quad j = 1, 2, \dots, m. \quad (21)$$

The corresponding number of the degrees of freedom is

$$\nu = 2(n - 1). \quad (22)$$

If the table for the Student distribution results in the absolute value $|t_{\bar{x}_k - \bar{x}_l}|$ that exceeds the value t_α determined on the basis of the Student distribution table, then one can conclude, with

the probability $Q \geq 1 - \alpha$, that the difference in the mean values \bar{x}_k and \bar{x}_l is insignificant and that the characteristics of interest are not different from the standpoint of their significance (influence), and can be put therefore in the same group. The next step is to evaluate the distributions of these characteristics from the factors affecting these characteristics.

3 | Application to a Teleoperating Task

3.1 | Description of the Task

In the context of a partnership with an industrial partner, a robotic platform has been built for the transportation of heavy loads, for example, the different parts of a plane, in different locations of a huge hangar. The robotic platform is teleoperated with a tablet by an operator standing close to the platform (see interface Figure 1). Using buttons, the operator can move forward or backward at low speed or turn left or right. As an assessment of the system was requested, a scenario has been defined for the test of the teleoperation, especially the different maneuvers, turning, bypassing obstacles, moving backward (see Figure 2). Starting in A, the operator has to move the robotic platform toward position B, avoiding a first obstacle. Then, he has to turn back going through a zigzag to reach position C and finally turn and move backward at the same time to reach the final position.

3.2 | Using the DM to Define the Assessment Criteria

The scenario was validated but a difficulty was to determine the criteria for the assessment. As different persons suggested different criteria, the DM has been used. Eight external experts familiar with human systems interactions had to be selected. Experts have been chosen among students following a master of science degree in human systems interactions (students of the Ecole Nationale Supérieure de Cognitique, France). Each of them has been asked to provide a list of criteria for the assessment. The main responses were: (a) Completion of the task Y/N, (b) Duration of the task, (c) Number of bumps into obstacles, (d) Feeling of the operator (questionnaire), (e) Number of trajectory corrections using the backward move, (f) Ergonomics (questionnaire). The list of all answers was provided to the experts to try to converge towards a shorter list. After the feedback, the experts provided new answers and they agreed upon three important criteria: (x_1) Duration of the task, (x_2) number of trajectory corrections, (x_3) feeling of the operator.

Importantly, if the list of criteria had been established according to a consensus based on majority (see Table 3), the completion of the task criterion would have been selected in the first place. However, after the first feedback (all responses were forwarded to everyone anonymously), one person suggested that there was no sense to test the task if it was not feasible and assumed that it was designed in such a way that its completion should not be an issue. The question was therefore forwarded to the technical team, who responded that it could indeed be designed and tested to make sure that it could be completed. As a consequence, the completion criterion has been rejected. This decision illustrates the benefit of

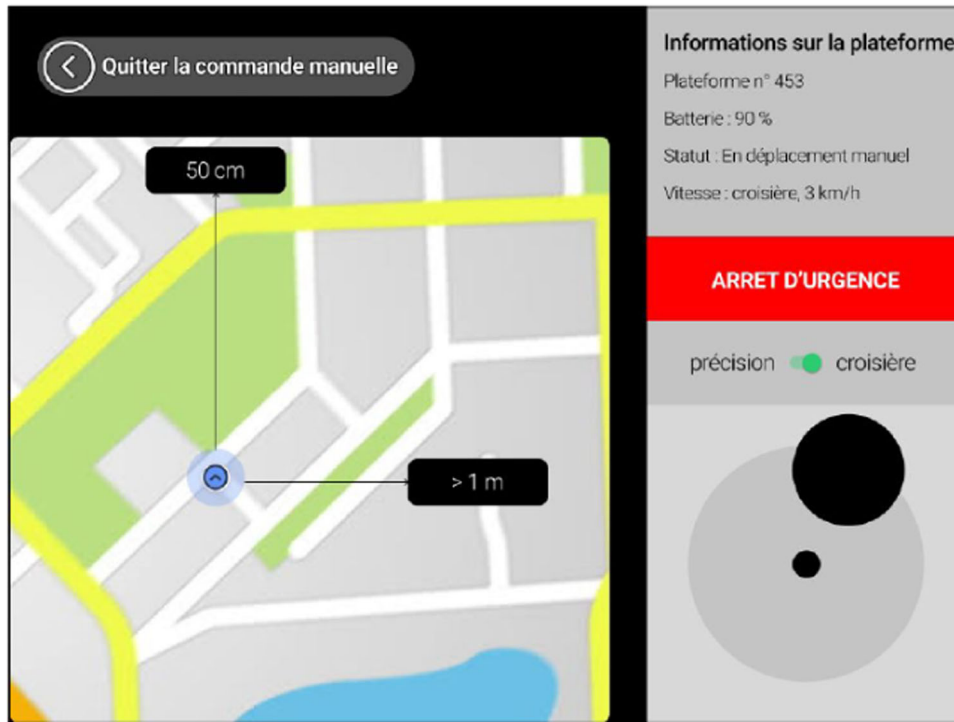


FIGURE 1 | Command interface for manual teleoperation (in French). [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/rob.22477)]

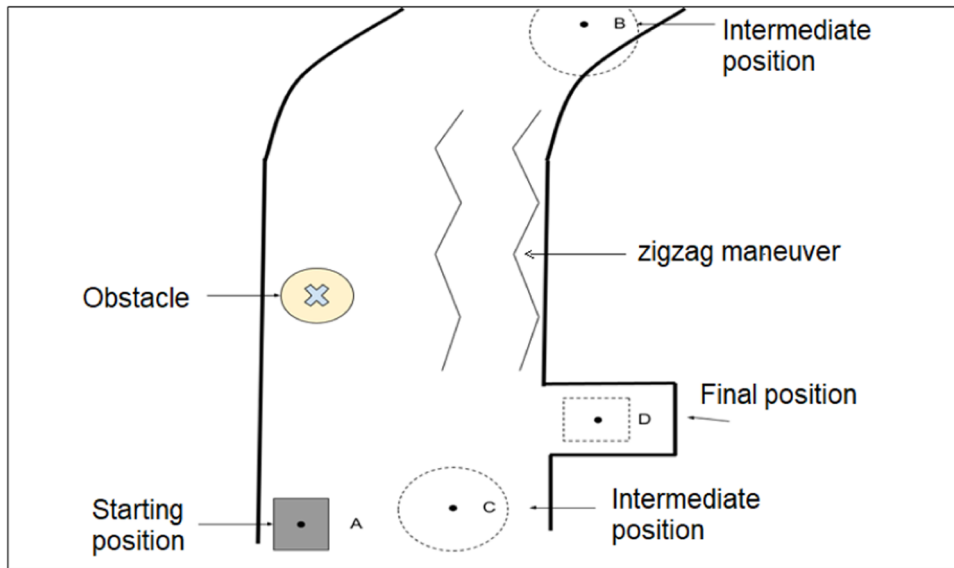


FIGURE 2 | Scenario for the assessment of the teleoperation task. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/rob.22477)]

the anonymous DM, which facilitates feedbacks from everyone, even if the person is the only one to think that way.

3.3 | Using the DM to Define the Weights of the Criteria

The objective is to calculate a global score based on the three variables x_1 , x_2 , and x_3 . The problem is that the variables are very different and cannot easily be combined. A possible approach is to convert each of them in a number between 0 and

1 and to use a weighted sum to get a global score. For the duration T of the task, considering an optimal path and the maximum speed, the minimum duration T_{\min} can be calculated. Then the following conversion is applied to obtain a number between 0 and 1:

$$\text{If } (T > 2T_{\min}) \text{ then } x_1 = 0, \text{ otherwise } x_1 = \left(-\frac{T}{T_{\min}} \right) + 2.$$

For the number of correction maneuvers x_2 , see Table 4 for the conversion function.

TABLE 3 | List of responses after first round.

Criteria expert	1	2	3	4
1	Completion of task	Duration	Number of corrections	N/A
2	Duration	Ergonomics	N/A	N/A
3	Completion of task	Duration	Number of bumps	Feeling of operator
4	Duration	Completion of task	Number of turns	N/A
5	Duration	Feeling of operator	Number of bumps	N/A
6	Duration	Number of corrections	Ergonomics	N/A
7	Completion of task	Duration	N/A	N/A
8	Duration	Number of corrections	Completion of task	N/A

Note: A selection based on majority votes would lead to (1) duration, (2) completion of task, and (3) number of corrections.

TABLE 4 | Conversion of the number of corrections in a number between 0 and 1.

Number of corrections	0	1	2	3	4	> 4
Conversion to get x_2	1	0.8	0.6	0.4	0.2	0

For the feeling of the operator, a questionnaire is used. The cognitive load, ergonomics of the interface, and the simplicity of the maneuvers are assessed with a number between 0 and 5, summed up, and finally linearly converted into a number between 0 and 1 to get x_3 .

The last problem is to determine the weight of each variable. The DM has been used. Each expert had to assign a rank between 1 and 3 to each variable, eventually with equality.

The obtained ranking matrix based on the experts' evaluations is shown in Table 5. The condition (5) is fulfilled. Indeed,

$$\sum_{j=1}^3 \sum_{i=1}^8 x_{ij} = \sum_{i=1}^8 \sum_{j=1}^3 x_{ij} = 48,$$

so that the Table 4 data are consistent. The calculated sums of the columns indicate that characteristic #1, with the sum $\sum x_{i1} = 11$ is the most important one and characteristic #3, with the sum $\sum x_{i2} = 23$ is the least important.

Since in the Table 4 matrix, some of the ranks are equal, Equation (8) should be used to evaluate the concordance coefficient. Equation (9) yields: $T_2 = T_3 = T_4 = T_5 = T_8 = 0$; $T_1 = T_6 = 6$; $T_7 = 24$; The correlation factor S computed in accordance with Equation (7), is $S = 78$. The concordance coefficient found on the basis of the Equation (8) is $C = 0.557$. To determine how far this coefficient is from zero we consider the F test. It yields: $F = 2.51$. The degrees of freedom are: $\nu_1 = 3 - 1 - \frac{2}{8} = 1.75 \approx 2$, $\nu_2 = (8 - 1) \times 6.833 \approx 48$. For $\alpha = 0.05$, with these degrees of freedom, we find: $F_\alpha = 3.2$. Since this value is greater than $F = 2.51$, we conclude, with the probability $Q \geq 0.95$, that there is a nonrandom agreement among the experts that the selected characteristics reflect well the value of the choice of interest.

Let us assess now, using Equations (15) and (16), the differences in the roles of different characteristics on the algorithms of interest and the significance of these roles.

As a normalization of the ranks has been made, the mean value of the ranks is 2. The Equations (15) and (16) yield: $D_1 = 4.875$; $D_r = 4.357$. The significance of the difference in the variances D_1 and D_r can be checked using Equation (17). It yields: $Z = \frac{1}{2} \ln\left(\frac{4.875}{4.357}\right) = 0.056$. With $\nu_1 = 2$ and $\nu_2 = 48$, for $\alpha = 0.05$, we find: $Z_\alpha = 3.2$. Since $Z < Z_\alpha$, we conclude, with the probability $Q \geq 0.95$, that the difference in the influence of the characteristics of interest is statistically not significant.

Determine now the influence of the role of the selected characteristics on the general evaluation of the choice under consideration. Different groups of the selected characteristics can be addressed based on the assessment of the difference in the mean values of the ranks of different criteria. The standard deviations of the criteria can be found, using the Equations (18), as follows: $\sqrt{D_1} = 0.44$; $\sqrt{D_2} = 0.38$; $\sqrt{D_3} = 0.35$;

The highest mean value of $\bar{x}_3 = 2.875$ has been given by the experts to the characteristic x_3 , followed by $x_2 = 1.75$. The Equations (20) and (21) yield: $D_{\bar{x}_3 - \bar{x}_2} = \frac{0.38^2 + 0.35^2}{8} = 0.033$, $\sqrt{D_{\bar{x}_3 - \bar{x}_2}} = 0.18$. Using the Equations (19) and (22), we have: $t_{\bar{x}_3 - \bar{x}_2} = \frac{2.875 - 1.75}{0.033} = 34.1$, $\nu = 2(8 - 1) = 14$. For $\alpha = 0.05$ and $\nu = 14$ the Student table yields: $t_\alpha = 1.76$. Since $t = 34.1$ is much greater than $t = 1.76$, we conclude, with the probability $Q \geq 0.95$, that the difference in the mean values of characteristics #3 and #2 is statistically significant, and therefore these characteristics should be placed, based on their roles, in different groups. A similar test can be performed for a comparison between x_1 and x_2 .

4 | Conclusion

The DM is suitable because collaborative robotics is multidisciplinary and it is often difficult to agree upon the relevance of different criteria, or to determine the weight of different parameters. In addition, it is well known that the Physical HRI and Cognitive HRI communities are addressing different questions and do not share the same expertise. As experts may have some

TABLE 5 | Normalized ranking matrix.

Variable Experts	x_1 Duration	x_2 No of corr.	x_3 Feeling	$\sum_{j=1}^3 x_{ij}$
1	1.5	1.5	3	6
2	2	1	3	6
3	1	2	3	6
4	1	2	3	6
5	1	2	3	6
6	1.5	1.5	3	6
7	2	2	2	6
8	1	2	3	6
$\sum_{i=1}^8 x_{i,j}$	11	14	23	48

difficulties to share the same view, the Delphi method, which allows anonymous expressions, is especially appropriate.

The DM, however, has some flaws. The main one is the weak interaction among experts. If the problem is very complex, long discussions are required among experts and if the anonymous method is chosen, too many rounds would be needed to reach a consensus, which is impractical.

The details of the three particular DM-related problems addressed in this analysis, namely, selection of DM experts, DM-based estimates using Student's distribution, and DM-based decision-making support using Fisher criterion, can be helpful when addressing the combination of different criteria for an assessment of a complex human–robot interaction.

We finally recommend further research on how to combine the DM and trust-based methods, for instance the weight penalty (Weidong et al. 2024), or the bidirectional feedback (Sun et al. 2023), depending on the type of problem addressed.

Data Availability Statement

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

- Adams, L. A. 1980. "Delphi Forecasting: Future Issues in Grievance Arbitration." *Technological Forecasting and Social Change* 18, no. 2: 151–160.
- Apraiz, A., G. Lasa, and M. Mazmela. 2023. "Evaluation of User Experience in Human–Robot Interaction: A Systematic Literature Review." *International Journal of Social Robotics* 15: 187–210. <https://doi.org/10.1007/s12369-022-00957-z>.
- Azani, H., and R. Khorramshahgol. 1990. "Analytic Delphi Method: A Strategic Decision Making Model Applied to Location Planning." *Engineering Cost and Production Economics* 20, no. 1: 23–28.
- Backley, C. 1995. "Delphi: A Methodology of Preferences More Than Predictions." *Library Management* 16, no. 7: 16–19.
- Bologaro, G. 1994. "Delphi Technique Can Work for New Product Development." *Marketing News* 28, no. 1: 11.
- Breiner, S., K. Cuhls, and H. Grupp. 1994. "Technology Foresight Using a Delphi Approach: A Japanese-German Cooperation." *R&D Management* 24, no. 2: 141–153.

Brockhaus, W. L., and J. F. Mickeldon. 1976. *The Delphi Method and Its Applications: A Bibliography*. Report of the Industrial College of the Armed Forces, 20319. Washington, DC.

Brown, B. B. 1968. *Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts*. Report of the RAND Corp. P-3925. Santa Monica, CA, USA.

Brown, T. 1972. "An Experiment in Probabilistic Forecasting," R-944-ARPA, The first RAND Paper.

Cambor, B., J. M. Salotti, C. Fage, and D. Daney. 2021. "Degraded Situation Awareness in a Robotic Workspace: Accident Report Analysis." *Theoretical Issues in Ergonomics Science* 23, no. 1: 60–79. <https://doi.org/10.1080/1463922X.2021.1879308>.

Coronado, E., T. Kiyokawa, G. A. G. Ricardez, I. G. Ramirez-Alpizar, G. Venture, and N. Yamanobe. 2022. "Evaluating Quality in Human-Robot Interaction: A Systematic Search and Classification of Performance and Human-Centered Factors, Measures and Metrics Towards an Industry 5.0." *Journal of Manufacturing Systems* 63: 392–410.

Dalkey, N. C. 1969. *Delphi Method: An Experimental Study of Group Opinion*. Report of the RAND Corp., RM-5388-PR. Santa Monica, CA, USA.

DeHaan, J., and R. Peters. 1993. "Technology: Toys or Tools? Results of a Dutch Delphi Study." *Information and Management* 25, no. 6: 283–289.

Demir, K. A., G. Döven, and B. Sezen. 2019. "Industry 5.0 and Human-Robot Co-Working." *Procedia Computer Science* 158: 688–695.

Goldfisher, K. 1992. "Modified Delphi: A Concept for a New Product Forecasting." *Journal of Business Forecasting* 11, no. 4: 10–11.

Goodrich, M. A., and A. C. Schultz. 2007. "Human–Robot Interaction: A Survey." *Foundations and Trends® in Human-Computer Interaction* 1, no. 3: 203–275.

Helmer, O. 1994. "Adversary Delphi." *Futures* 26, no. 1: 79–87.

Koynak, E., J. Bloom, and M. Leibold. 1994. "Using the Delphi Technique to Predict Future Tourism Potential." *Marketing Intelligence and Planning* 12, no. 7: 18–27.

Lazarus, J. V., D. Romero, C. J. Kopka, et al. 2022. "A Multinational Delphi Consensus to End the COVID-19 Public Health Threat." *Nature* 611, no. 7935: 332–345. <https://doi.org/10.1038/s41586-022-05398-2>.

Linstone, H. A., and M. Turoff. 2002. *Delphi Bibliography*. <http://is.njit.edu/pubs/delphibook/appendix.pdf>.

Mitchell, V. W. 1992. "Using Delphi to Forecast in New Technology Industries." *Marketing Intelligence & Planning* 10, no. 2: 4–9.

Moulières-Seban, T., D. Bitonneau, J. M. Salotti, J. F. Thibault, and B. Claverie. 2016. "Human Factors Issues for the Design of a Robotic System." In *Proceedings of the AHFE 2016 International Conference on*

Human Factors in Robots and Unmanned Systems, Walt Disney World, Florida, USA, July 27–31.

Mutlu, B., and N. Roy. 2016. “Cognitive Human-Robot Interaction, Chapter 71.” In *Springer Handbook of Robotics*, edited by B. Siciliano, and O. Khatib, 1835–1874. Berlin: Springer.

Sheridan, T. B. 2016, “Human-Robot Interaction: Status and Challenges.” *Human Factors* 58, no. 4: 525–532. <https://doi.org/10.1177/0018720816644364>.

Suganthi, L., and T. R. Jagadeesan. 1992. “Energy Substitution Methodology for Optimum Demand Variation Using Delphi Technique.” *International Journal of Energy Research* 16, no. 9: 917–928.

Suhir, E. 1996. *Viability Audit of Engineering Projects: Application of Delphi Technique*. Report 11174. New Jersey, USA: Bell Laboratories Department.

Suhir, E. 1997. *Applied Probability for Engineers and Scientists*. New York, NY, USA: The McGraw-Hill Companies, Inc.

Suhir, E. 2011. “Human-in-the-Loop: A Way to Assess the Role of the Human Factor In Vehicular Mission-Safety-and-Success Problems.” In *IEEE/AIAA Aerospace Conference*, Big Sky, Montana, March 5-12.

Suhir, E. 2018. *Human-in-the-Loop: Probabilistic Modeling of an Aerospace Mission Outcome*. 1st ed. Boca Raton: CRC Press, Taylor and Francis Group. <https://doi.org/10.1201/9781351132510>.

Sun, Q., F. Chiclana, J. Wu, Y. Liu, C. Liang, and E. Herrera-Viedma. 2023. “Weight Penalty Mechanism for Noncooperative Behavior in Large-Scale Group Decision Making With Unbalanced Linguistic Term Sets.” *IEEE Transactions on Fuzzy Systems* 31: 3507–3521.

Villani, V., F. Pini, F. Leali, and C. Secchi. 2018. “Survey on Human-Robot Collaboration in Industrial Settings: Safety, Intuitive Interfaces and Applications.” *Mechatronics* 55: 248–266.

Weidong, J., G. Tiantian, C. Mingshuo, Z. Mi, and W. Jian. 2024. “A Personalized Bidirectional Feedback Mechanism by Combining Cooperation and Trust to Improve Group Consensus in Social Network.” *Computers & Industrial Engineering* 188, no. C. <https://doi.org/10.1016/j.cie.2024.109888>.