

Title: The use of pigs vocalisation structure to assess the quality of human-pig relationship

## Letters to reviewers and Editor following Round #1

Comment of Editor:

Status: Moderate revisions

This manuscript is of high interest. It however deserves amendments before we can recommend it. The two reviewers made detailed and complementary comments that should help the authors.

The writing needs to be improved to help the reading. The English needs also editing by a native english speaker.

Dear editor,

We now have made a revision of our manuscript. We answered all comments from both reviewers and the english was edited for the entire manuscript. We hope our changes will meet the reviewers' expectations.

See below the detailed answers to each reviewer.

All the best

Avelyne Villain and Céline Tallet, corresponding authors

### Reviews

***Reviewed by Matteo Chincarini, 06 Apr 2022 07:44***

Dear Editor,

The manuscript "The use of pigs vocalisation structure to assess the quality of human-pig relationship" is addressing an original work on interspecific acoustic communication and explores the non-invasive emotional indicators in swine. The study design is very accurate, and the discussion of the results gives an exhaustive overview of the topic. A couple of concerns could be related to my misunderstanding. The first one regards the pen size where the experiment has been conducted and the second one is on the assumption test for PCA (please, see below). The experiment is well described and even if, working with farm animals and acoustic analysis is very challenging, the authors have worked very hard to set up elegant research. Furthermore, there is growing interest in vocal communication related to animal welfare as well as human-animal relationships. Finally, these results provide potential non-invasive indicators relevant to animal welfare. The manuscript needs also some minor revisions relative to the figures.

Below are some specific comments to the authors for minor revisions:

Dear reviewer,

Thank you for your review and your enthusiasm on the manuscript. We have addressed your comments and we hope our responses will clarify the concerns you had on the previous version of the manuscript. Please find below the detail answers (in blue) to your comments, with line referencing to the new version of the manuscript.

Title

The title clearly reflects the content of the article.

## Abstract

The abstract is concise and presents the main findings of the study. I've only one concern regarding the first sentence: "In domestic species, studying human-animal interactions and their consequences on the establishment of a positive Human-Animal Relationship (HAR) would have applications for both improving animal welfare." I'm not a native English speaker but here maybe it is possible to rephrase. I understand that authors are referring to animals and humans when they use the term "both" but now it seems it is referring to "interactions" and "consequences".

The sentence has been changed. "Studying human-animal interactions in domestic species and how they affect the establishment of a positive Human-Animal Relationship (HAR) may help us improve animal welfare and better understand the evolution of interspecific interactions associated with the domestication process." line 18.

## Introduction

Hypotheses have been explicit very clearly and they are supported by several papers representing the state of the art in this field.

Thank you for this comment.

LL 46-47: please consider merging these two very short sentences.

Former version "Domestic species form particular relationship with humans. In farms, this relationship is important for animal welfare."

This was changed to "In farms, the relationship that domestic animals form with humans is important for animal welfare" line 52

## Materials and methods

This section is, in general, well explained and detailed.

Thank you for this comment.

## Ethical note

L 103: please, consider citing the French and European legislation (this will be relevant for the pen size, see below)

This text was added: "UE3P, where the experiment was carried out, is an experimental unit authorized by the French Ministry of Agriculture to breed animals for experimentation under the number D35-275-32. This authorization includes a derogation to follow the directive 2008/120/EC relative to the protection of pigs and its regulations." line 113

## Subject and housing conditions

L 112: please, verify the pen size according to your authorisation. According to the EU Directive 63/2010, the minimum enclosure size should be 2,0 m<sup>2</sup> (Table 7.3). This could be not your case but it needs to be justified (it can be smaller due to experimental grounds).

You are right, but the experimental facilities have a derogation, as INRAE's experimentation are done to develop application for livestock breeding. See response above and additional sentence in the manuscript "UE3P, where is experiments were carried out, is an experimental unit authorized by the French Ministry of Agriculture to breed animals for experimentation under the number D35-

275-32. This authorization includes a derogation to follow the directive 2008/120/EC relative to the protection of pigs and its regulations.” line 113

### Conditioning

L 137: you say that conditioning took place between day 42 and 62 after weaning, so it would be between 70 and 90 days of life? At L 107 it is reported that piglets involved in the study were from 28 to 62 days of life, please double check it.

Thank you for this careful read, “day” refers days of life and this was made clear. This was a typo as the beginning of the experiment was after weaning (at 28 days of age).

“From day 28 (day of weaning) to day 39 of life, pigs were separated into two groups that experienced a different post-weaning period as follows:” See line 131

L 145: I’m not sure what “Hens” means here

Multiple typos were corrected, it was one of them.

### Behavioural monitoring and analysis

Please, specify if the behaviour has been analysed either by the same or different persons.

Yes, only one person scored the videos for behavioural analyses. See line 180 “ For every second trial, the two-minute reunions with the human were analysed by the same person : trials number 2, 4, 6, 8, 10 and 11”

### Acoustic monitoring and analysis

L 178: Even if Praat is well-known software in this field, I think it would be better to cite it using a reference. Especially to be clear about the version that has been used (please, you can give a look here: [https://www.fon.hum.uva.nl/praat/manual/FAQ\\_\\_How\\_to\\_cite\\_Praat.html](https://www.fon.hum.uva.nl/praat/manual/FAQ__How_to_cite_Praat.html))

Yes. See line 210 “Praat software (Boersma and Paul 2001), version 6.0 from <http://www.praat.org/>.”

### Statistical analysis

I am not a statistician. However, when applying PCA I think it is worth reporting the value of Kaiser-Meyer Olkin (KMO) and Bartlett’s test of sphericity as preliminary tests (or explain why not). Maybe, you could also consider using the Measure of Sampling Adequacy (MSA) if some variable needs to be excluded.

From what we understood of the literature on the subject, Kaiser-Meyer-Olkin test is a measure of how the data are suited for Factor Analysis (<https://www.statisticshowto.com/kaiser-meyer-olkin/>). Although both a Factor Analysis and a Principal Component Analysis identify patterns and correlations between variables, they do not rely on the same assumptions. Contrary to the Factor Analysis, the mathematics behind the PCA does not assume the existence of latent factors underlying the observed data (<https://www.displayr.com/factor-analysis-and-principal-component-analysis-a-simple-explanation/>). In our case, the PCA is used to build composite scores (either composite behavioural scores or composite acoustic scores) to reduce the number of statistical variables (and avoid type I errors testing each variable one after the other). In addition, when measuring several vocal parameters, it happens indeed, that some of them are correlated and thus

load strongly together (see for example [Briefer et al 2019 on pig grunts](#)). The PCA thus also allows us to visualize the parameters that load together on PCs and gives us a rationale for understanding the global acoustic structure of the calls (and not for clustering purposes for example). As a consequence, in our case, neither the Kaiser-Meyer-Olkin test nor the Bartlett's test are necessary.

From Table 4, I'm understanding that you transformed some variables (like using log or sqrt). Please, consider adding this information also in the text when you write about symmetrical distribution L 206.

Yes, see line 245: " linear transformations were computed when necessary to reach symmetrical distribution (see tables 2, 3, 4)."

L 207: "pca", did you mean function "dudi.pca"?

Yes, see line 248: "'dudi.pca' function from 'ade4' R package (Dray and Dufour 2007, 4)"

L 254: inside the code, I think ID/time/Phase should be ID/Time/Phase

Yes, see line 301 "Model2 <- lmer ( Vocal variable ~ Treatment \* Phase \* Time + Treatment \* HumanID + Time \* HumanID + Treatment \* Replicate + Time \* Replicate + (1 | PigID/Time/Phase) , data= dataVocallsolation + dataVocalReunion)."

## Results

General: most of the script has been reported, could you add also the PCA analysis?

Regarding data availability:

- we have shared all datasets used in the study
  - we have written a readme to guide readers through the dataset and explained which dataset corresponds to which analysis (<https://doi.org/10.15454/RTBO3O>).
  - in the manuscript, we have made sure to report which R libraries and which functions in these libraries we used. All formulas of the statistical models are explicit in the text to facilitate transfer of information and replicate the analysis. All libraries are open source as well. See statement in the manuscript: "We have made sure to report in the main text of the article which R libraries and which functions in these libraries we used. All formulas of the statistical models are explicit in the text to facilitate transfer of information and replicate the analysis. All libraries are open source as well."
- Line 741
- the PCAs were performed on raw parameters contained in the dataset we shared. All pre-processing transformations on parameters are reported in the manuscript. So the PCAs can also be redone from the datasets we shared.

This way anybody can redo any of the analysis represented in the paper.

L 344: figure 3 is not present in the manuscript

The problem of figure referencing was solved.

L 369: figure 4, I'm not sure that is referring to the actual figure

The problem of figure referencing was solved.

Tables and figures

Please, see above

Corrected.

Discussion

The discussion is exhaustive and well supported by the literature. Still, the conclusions are not overstated.

Thank you for this comment.

L 475: there are two “first”

Corrected.

References

Fine.

***Reviewed by anonymous reviewer, 04 May 2022 14:19***

Review of “The use of pigs vocalisation structure to assess the quality of human-pig relationship” by Villain et al.

This is an interesting study aiming to investigate if (changes in) pig vocalisation reflects the quality of human-animal relationships. I have, however, some major concerns and a few other issues, which I have summarised below.

Dear reviewer,

We thank you for your careful read of the manuscript and we apologize if the spelling made it difficult. We appreciated your comments and we did our best to answer to all of them, providing changes in the manuscript and adding tables as supplementary to meet your expectations. Please find below the detail answer to your comments (in blue), with line referencing to the new version of the manuscript.

Major concerns:

The study sets out to analyse pig vocalisation in different situations, comparing changes in sound structure both within and between pigs when treated differently over time. However, the data are immediately reduced to only grunts (line 178) because they were the most frequent. Is frequency important for all calls? One scream may say more than a thousand grunts, to paraphrase Ibsen. I would like to see a couple of phrases explaining this a bit more, as this is an important aspect of your data editing. You also focus on vocal quality, but what about vocal quantity? You mention vocal activity in line97, but refer to some qualitative aspects there, too.

We added a table in the supplementary material collecting the number of each call type per test and per treatment as well as the number of pigs involved in the count.

We hope this table will convince that regarding the number and the experimental design, an analysis of the quality of the vocalizations produced was not possible. Page 22 of the supplementary material and reference to this table in main text line 218.

Table S5: Number of calls of each call type recorded during the session and the number of pigs involved in the count. Taking into account the different statistical variable that needed to be add in the models, and thus the number of calls and pigs needed to have reliable statistical analysis, it was thus decided to use only grunts in this study.

Call type	Treatment	Before conditioning (Isolation/Reunion test – Reunion with H)		During conditioning (all trials pooled)		After conditioning (Isolation/Reunion test – Reunion with H)	
		N calls	N pigs	N calls	N pigs	N calls	N pigs
bark	H	14	5	13	7	6	3
grunt	H	670	21	3979	29	1981	25
mixed	H	8	1	172	12	157	9
scream	H	0	0	14	4	39	2
squeal	H	11	2	94	10	66	11
bark	H+	4	2	18	6	1	1
grunt	H+	1244	27	5006	29	2072	27
mixed	H+	0	0	142	12	21	3
scream	H+	0	0	7	2	0	0
squeal	H+	8	2	50	8	25	6

I am missing which specific hypotheses you are testing? Or rather, in the Discussion, you dismiss one of the hypotheses, and then appear to suggest two new ones (lines 539-544).

In the paragraph before, we discuss the effect of the mere presence of a human on the structure of grunts, leading to two hypotheses. The aim was to announce these hypotheses and address them in the next paragraph “the interpretation of the second type of human-piglet interactions may allow to address these hypotheses” (in the version 1) but this was not clear enough.

We worked on the writing to increase clarity on that aspect. Lines XXX to line XXX:

Beginning of discussion section:

- “In the next two paragraphs, we discuss the results of the standard reunion test before and after conditioning. This enables us to raise two possible hypotheses. We then use the results of the sessions of additional positive contacts of the conditioning to discuss these hypotheses.” Line 583

End of paragraph of discussion of Isolation/Reunion test

“This test may allow us to suggest two potential non exclusive hypotheses to explain why the effect of human proximity on grunt acoustic structure attenuates as the familiarity to the human increases. In a first hypothesis, we could think that this attenuation of acoustic flexibility is due to a decrease in reactivity to the human, which may be linked to a disinterest of human contacts and an increase in foraging natural behaviours. In another hypothesis, this attenuation of acoustic flexibility may be due to a violation of piglets expectations: because the human remains static during the test, this may inhibit vocal reactions to the proximity. The interpretation of the second type of human-piglet interactions below may allow to address these hypotheses.” Line 638

Was the vocalisation when conspecifics were social partners (line 141) used at all? I may have missed where that is presented. And if not, why was this included? And if used, how did you separate the vocalisation of the different pigs?

Vocalisations in relation to conspecifics arrival and the experimental design of the conditioning is already published<sup>1</sup>. The reason why the two-way conditioning is explained in the method (with the human as the outcome and with the penmates as the outcome) is to be perfectly transparent on the full experiment and the different (pseudo)social experiences the experimental pigs were subjected to. The present article complete the preceding one on the same design, focusing on human-pig relationship.

Indeed, for ethical purposes, to limit the number of animals bred for experimental purposes, one experiment was designed with two (compatible) objectives. 1) Study the vocal and behavioural anticipation of (pseudo)social partners [using data of the conditioning before the reunion occurred, see Villain et al 2020, scientific reports]. 2) Study vocal and behavioural evolution of human-pig interactions [using data that were not explored in study 1]

See paragraph in the revised version of the manuscript Line 172: "Sessions of reunions with social partners were not studied and only served as reward during the conditioning in a previous analysis of vocal expression of positive anticipation (Villain et al. 2020). Indeed, first the two (pseudo) social contexts would have been difficult to compare (reunion between three pigs vs. reunion between one pig and one human). Second, regarding the vocal behaviour, the caller among the group of three pigs would not have been identified reliably, making it difficult to study within individual vocal flexibility"

The analyses are complex and can be difficult to follow in places. Is a p-value threshold of 0.05 too large for 3-way interactions? Some (many?) of your 3-way interactions have a p-value of 0.03 (even 0.07, which you still keep?), and I am left questioning how relevant they are. It leads to results like "grunts produced closer to the human were shorter... but only in untamed piglets, effect being stronger before the conditioning" and "grunts had a higher frequency range ... when produced closer to the human..., but only in untamed piglets and before the conditioning". On a data set of this size, I wonder to what extent these results can be generalised. It makes the manuscript very long and very difficult to follow in places – and the main results drown. Is there enough power to make such detailed conclusions?

This experiment has only a relatively complex design: two independent groups of pigs (H and H+) are subjected to a conditioning and their behaviour is studied before, during and after the conditioning: the purpose is clearly to study time effect in interaction with the treatment positive handling at weaning. However, several reports in the literature made us add some factors that we thought relevant regarding our question. Following previous reports of effects of the spatial proximity to a human on pig vocal behaviour, the 'location' of the pig also needed to be taken into account, especially because the proximity of the pig is a relevant descriptor of human-pig interactions. We thus used the three way interaction between "treatment\*conditioning time\*proximity" and the large sample size in the dataset allowed to do so.

Regarding your question on p value threshold, from our understanding of statistical models and testings, increasing the degree of interaction just makes the p-value threshold of 0.05 harder to reach but we don't recall papers in the literature suggesting to change the significant threshold when dealing with three-way interactions. If we missed something, we would be happy to try a different model and/or a different significant threshold if suggested in the literature.

---

1 Villain et al., 2020 "Piglets Vocally Express the Anticipation of Pseudo-Social Contexts in Their Grunts."

The outcome of the statistical results may thus look technical, but unfortunately, we think we must take the architecture of the experiment into account and that we must not oversimplify models and risk having conclusions beyond support.

Concerning the p-value between 0.05 and 0.1. Indeed, we did report one (problematic) three-way interaction: line 348 of the **version 1 of the manuscript** "For AcPC2, the three-way interaction was close to reach significant level ( $\chi^2_1 = 3.3$ ,  $p = 0.07$ ), thus, for conservative purposes, the results of the post hoc tests of the three-way interaction are presented (see two-way subsequent interactions comparisons in supplementary tables S2 and S3)". Another sentence also explained this in the supplementary material "Note : due to a three-way interaction close to significance level, contrasts were generated with the three-way interaction and with the two-ways interactions of interests". Following your comments, we decided to keep the supplementary material as it was but to change the main text showing results of post hoc test on the two-way interactions (and not the tendency on the three-way interaction). However, we decided to keep a sentence of this three-way interaction to advertise the reader that this analysis also exists in the supplementary material if needed. The text has been changed

"For VocPC2, the three way interaction did not reach significance ( $\chi^2_1 = 3.3$ ,  $p = 0.07$ ), so only subsequent two way interactions were considered (but post hoc tests on the three way interaction can be found in supplementary, tables S1 to S3). For VocPC2, significant two way interactions were found between the conditioning time and the location ( $\chi^2_1 = 10.3$ ,  $p = 0.001$ ) on the one hand, and between the location and the treatment ( $\chi^2_1 = 4.2$ ,  $p = 0.04$ ) on the other hand. Post hoc tests revealed that grunts produced closer to the human had a higher VocPC2, meaning they had a higher pitch, effect being stronger before the conditioning than after (before: away – close, z.ratio = -6.12,  $p < 0.001$ ; after: away – close, z.ratio = -2.88,  $p = 0.004$ , figure 3C). The increase in VocPC2 with the location was greater for non handled piglets than positively handled piglets (H piglets: away – close, z.ratio = -5.54,  $p < 0.001$ ; H+ piglets: away – close, z.ratio = -3.82,  $p = 0.001$ , figure 3D)." Line 412

See also changes in figure 3, line 397.

Our results are in line with the literature on the effect of positive handling on behaviour and vocal quality, and our statistical approach reasonable, and thus their generalisation makes no real doubt to us. However, for sure it cannot be proven without any replicate of the design and will remain questionable.

Line 445-449: Is this what you expected? Could your interpretation be affected by the nature of the treatment, in other words, you will describe the response of the H+ pigs as positive (either being touch/attention satiated, or know they can always come back, and therefore disinterested, or confident enough to go exploring). How can this be disproven if you haven't set out expected outcomes from the start?

Quote from version 1 of manuscript: "In addition, tamed piglets expressed more exploratory behaviours than untamed piglets after the conditioning (ReuPC3), which may be interpreted as natural foraging and disinterest from human contact, which may be a sign of positive welfare (Weerd and Day 2009). However, this could be interpreted also in terms of attachment to the human"

Expression of natural foraging has been hypothesized in previous literature on animal welfare and we cite this literature in the introduction "the decrease of experiencing negatively perceived contexts and the increase in experiencing positively perceived contexts and species-specific behaviors (Peterson, Simonsen, and Lawson 1995; Weerd and Day 2009)." Line 57.



In the discussion we thus confront our result of increased investigation of the room in more familiarised piglets (H+) to this literature, along with other marker of positive perception (attraction, contact to the human and time spent in proximity). All piglets at the end of the conditioning expressed behaviours in favour of an interpretation of positive perception of the human and H+ piglet expressed “something more”. We thus hypothesis later a sequential establishment of a positive HAR in piglets : attraction first and expression of foraging behaviour in a further “step” of relationship .

Paragraph of the discussion has been rephrased:

“This test also showed that the conditioning modified the behaviour of non handled piglets so that they finally expressed a similar attraction toward the human as positively handled piglets, after the conditioning. These results are in line with the behavioural results of the sessions of additional positive contacts. The analysis of piglets’ behaviour every second sessions of the conditioning showed that, although positively handled and non handled piglets started with different degree of proximity toward the human (trials 2 and 4, CondPC1), then, over time and for both treatments (H and H+), piglets expressed a higher attraction toward the human (CondPC1) and avoided less the human when the latter attempted to interact with them. At the end of the conditioning, piglets from both groups had similar level of proximity toward the human (trials 8, 10, 11 CondPC1)” Line 540

And:

“Beside behavioural proximity, piglets that were positively handled at weaning expressed more exploratory behaviours than non handled piglets after the conditioning (ReuPC3). This was also observed during the sessions of additional positive contacts of the conditioning: positive handled piglets started with a higher score associated with investigation than non handled piglets (CondPC2) and it held over the conditioning. Piglets that were positively handled at weaning also expressed a higher mobility than non handled piglets (CondPC3). These observations may be interpreted as an expression of natural foraging and disinterest from human contact, which may be a sign of positive welfare (Weerd and Day 2009). In addition, this could also be interpreted in terms of attachment to the human. Indeed, attachment to a human may facilitate exploration of novel environments or objects, as shown in dogs (Palmer and Custance 2008). A period of positive handling at weaning may provide an environment secure enough for the piglets to explore their environment in the presence of the human. Attachment has also been hypothesised in the lambs-human relationship (Tallet, Veissier, and Boivin 2009).” Line 556

And:

“We may be able to hypothesize a sequential establishment of a positive HAR over time: firstly with a decrease of attentive state and an increase in proximity and accepted contacts, and secondly with a disinterest of human contacts and the expression of natural foraging behaviour. The latter may require a higher exposure time.” Line 573

The last major concern is the text. The manuscript has a large number of typos (e.g. lines 405, 440, 463, 552 and elsewhere) and missing spaces, which leaves the impression that the uploading was rushed. I am usually quite forgiving when the English is a little bit rustic when the authors are writing in a second language. Unfortunately, there are places where it makes it difficult to understand what is meant (e.g. line 38: “a carrying human”; line 112: “on plastic duckboard and panels visually isolated pens”; line 568: “‘AH’ was more entitled to trigger higher positive states”), and as a reviewer, it can be jarring to have to second-guess the content. I would therefore recommend that the revised manuscript is copy-edited by third party before re-submission.

The manuscript has been proof read by a native English speaker.

Other issues:

The use of the word 'taming' is confusing, as domestic pigs (*Sus scrofa domesticus*) are all tame compared to the wild boar (*Sus scrofa*). I suggest to change this to 'positive handling'.

It was changed in the version 2 of the manuscript.

The authors refer to their study animals as piglets, but they are weaned and thus should be called pigs.

We agree it depends on the articles. Nevertheless, since one part of the experiment is already published (anticipation of rewards) and the term "piglets" has been accepted in the paper, we decided to keep piglet in this version to keep it homogeneous.

Line 19: objectify the quality?

Changed to "Understanding and describing". Line 21

Line 20: Is not all behaviour spatial?

Spatial behaviours, opposed to postural behaviours that could not track in the study (ear and tail posture, facial expression). Changed to "social, spatial and postural behaviours" Line 22.

Line 25: replace breeding with husbandry

Changed to "husbandry" Line 27.

Line 50: Animal welfare conveys? Consists of?

Wording was changed.

Line 69 and elsewhere: associated with, not to

This was corrected everywhere needed.

Line 88: What is 'formant'?

Definition added in parenthesis. See line 94 "as well as higher formants (which are frequency peaks containing more energy than others)"

Line 113: metal chain?

Yes.

Line 145 and elsewhere: Hence

Yes.

Figure 1. Suggest using 'area' for the distal and proximal areas, to not confuse it with the 16 zones. It is not clear where the distal area is (above or below the dashed line).

It was clarified in the legend of the figure and the term "area" was used everywhere it was needed in the manuscript. Thank you.

Line 161: Where the sessions recorded or was the annotation done live?

Only videos were used. It was clarified. See line 193 "Sessions and tests were recorded using a camera (Bosh, Box 960H-CDD) and behaviours were scored *a posteriori* on videos using *The Observer XT 14.0* (Noldus, The Netherlands) software."

Line 170: Replace over with other

Yes. Done.

Table 1: "The number of times the piglet looked at other parts of the room" - as the pig will be looking somewhere at all times, this will always be within 1 of the previous variable (Nb looks toward human).

This code was to distinguish when the pig has the head down from when the pig has the head up but not watching the human (watching doors or walls). The description was changed to "The number of times the piglet looked at other parts of the room than the human or the floor (walls, doors)" Line 205.

Table 2: The Table does not show "Behavioural response score for the reunion phase of the Isolation/Reunion test." but "Percentage of explained variance and variable loadings of the principal component analysis for the first three PCs."

Legend of table 2 and table 3 were rephrased

*"Table 2: Percentage of explained variance and relative loadings of parameters on PCs, following the Principal Component Analysis computed on the behaviours scored during the reunion of the Isolation/Reunion test. The first three PCs, having an eigenvalue above 1, constituted three behavioural scores: ReuPC1, ReuPC2, ReuPC3. Parameters that explain the most each PC are bolded (|loading|>0.4)." Line 256*

*"Table 3: Percentage of explained variance and relative loadings of parameters on PCs, following the Principal Component Analysis computed on the behaviours scored during the sessions of additional positive contacts of the conditioning. The first three PCs, having an eigenvalue above 1 constituted three behavioural scores: CondPC1, CondPC2, CondPC3. Parameters that explain the most each PC are bolded (|loading|>0.4)." Line 262*

Table 2 and elsewhere: You use the word 'parameter' when 'variable' is the correct term.

In practice, parameters were used to build composite scores, used as response variables in statistical model. So, we tried to be consistent using the term "parameters" for specific measures (a behaviour or an acoustic parameter), the term "score" to refer to the PCs and the term "variable" for statistics. We doubled checked the consistency throughout the manuscript

Table 4: Are these variables on vocalisation characteristics essentially showing the same, leading to a high loading for all of them on PC1, i.e. so highly correlated that they are not all needed?

Yes, several parameters measured on vocalization, sometimes load on the same PCs and are sometimes highly correlated. It is one of the purpose of using a PCA, to quantify which parameters load together. Depending on the study, not the same parameters will load on PC1 for example,

sometimes parameters describing the noise components will load on a different PC than the one describing the frequency distribution (see Briefer et al 2019 on pig grunts for example). Nevertheless, the PCA is used to have a non biased description of the structure of the vocalization, maximizing the variance, without having to pre select parameters.

Line 260: Vocal response variable

See response above

Lines 323 and 326: Why are these estimates?

Since no figures of this specific analysis is presented in the main text to see the range of variation depending on contexts (not the main scope of the study but necessary as a control), we report the output of the model and thus the estimate and 95% confidence interval along with the statistics.

Figure 4: It is not possible to know what comparisons the letters refer to, as some of them have no letter. What is different from what? Question if a three-way interaction on a subset of data is biological relevant? The blue and grey colours are indistinguishable.

The contrast of the blue and grey scales was increased. We wrote "subset" of data to explain that the Isolation period of the test was not included in this analysis. Indeed, the isolation phase was just used as a negative control but the main scope was to analyse the sessions of reunion with the human. So the term subset may be misleading. The "subset" here constitutes an entire dataset of all grunts produced during the reunions with the human. The term was thus removed. See our response above regarding the the use of the three-way interaction. When no letters are used, it means that the groups does not significantly differ, a sentence was added in the legend.

*"Figure 3: Acoustic structure of grunt during the reunions with a silent and static human (Isolation/Reunion test). Effect of conditioning (before or after), treatment (H or H+), and location of the pig relatively to the human (close: dark blue or away from them: light blue). Violin plots representing the median and the density of data distribution in the considered groups. (A, B) Results of post hoc tests following the significant three way interaction between the treatment, the conditioning time and the location on grunt duration (A) and on the first vocal score (-VocPC1, B). (C,D) Results of post hoc tests following the significant two way interactions between conditioning time and location (C) and between treatment and location (D) on the second vocal score (VocPC2). When involved in interaction, the conditioning time was fixed (as it was relevant to consider difference affected only by time). It thus allowed pairwise comparisons of interacting location and treatment (A, B) or levels of location (C). Letters represent significantly different groups ( $p < 0.05$ ). When no letters are present, no significant difference between groups was found. Stars (\*) between two groups represent a statistical trend ( $p < 0.10$ ). Full statistical report is available as supplementary material (tables S1 S2 for statistical test and S3 for model estimates)." Line 395.*

Figure 5: What does H:N mean in the legend of A? Figure D is missing (referred to in Table heading). Not sure what C means - failed used in figure heading, missed used in y-axis label, but what does this show?

"N" was used to refer to the trial number and as the stars next to N was referring to the significant effect of Trial number. It was removed. The entire legend was updated.

*"Figure 5: Evolution of vocal scores over the conditioning, during the 2min sessions of additional positive contacts. (A, B) Violin plots representing the median and the density of data distribution in the group. Interacting effect of location (in proximal area of the human ('close': dark blue) or elsewhere in the room ('away' from the human: light blue) and treatment (H vs. H+ pigs) on grunt duration (A) and VocPC2 (B). (C) Mean  $\pm$  SE per group, interacting effect of trial number and location of pigs on VocPC2. Different letters in A and B represent significantly different groups, "\*" in C represents significant*

*difference between the two slopes. Full statistical report is available as supplementary material (tables S1-S3).*" Line 457

Line 369: Here you refer to the wrong Figure. This was carried over for the rest of the manuscript, so that wrong Figures were indicated.

Yes, the problem has been solved.

Figure 6: Are these slopes based on linear regressions? is this justified?

On the figure, the raw data mean +/- se are indicated. Since the model was linear and the trial number was continuous, indeed, the estimates of the model are based on linear regressions (see slope estimates in statistical table S3 of the supplementary material)

Table 5: Does this not indicate either that vocal parameters are not very robust measures, or that you were unable to standardise your treatment?

We discuss that point in the discussion. These effects of the identity of the human needed to be reported and we thus suggest interpretations and future work in the discussion regarding this. Experimenters either failed mimicking each other (see line XXX, where it is specified they did their best)

"The experimenters tried to imitate each others behaviours (remote video monitoring) to decrease variability." Line 147

"In our study, both humans that interacted with the pigs wear exactly the same clothes and standardized their tactile interactions toward the pigs before starting the study, and agreed on the rhythm and types of sounds (words, intonation) to use, to minimise generating variability although no systematic controls of the human behaviour or spectral feature of voices were performed here" line 697

Differences between the humans, like odors, may explain (like you suggest later and we added that point to the discussion as well.

"Our results show that the identity of the human may modulate piglet proximity and vocal behaviour but the design of this experiment does not allow to find the causes of these observations (behaviour, voice characteristics, or even odour profile)." Line 703

This could explain why the two experimenters had different effects on behavioural proximity of the pigs and vocal scores. An analysis of the experimenter's behaviour may add information to disentangle these points but we think it is out of the scope of this paper and we suggest future work. Since both behavioural scores and vocal scores were affected (and not only vocal scores) we hypothesize that characteristics of the human may impact the effectiveness of positive handling, rather than the robustness of measures. But again, these effects bring new questions to the field.

Line 443: attraction instead of attractiveness

Yes, corrected there and elsewhere.

Line 467-468: This is using your interpretation of the positively handled pigs to draw conclusions on the limitations on the control group. The set-up did allow them to explore.

We are not sure we understood your point. We are discussing the fact that one possible "model" in terms of establishment of HAR could be that first the fear is reduced and the attraction is increased but that the disinterest of human contact and natural behaviours may come later in the process. Please see the new version of the paragraph to see if we understood your point.

We may be able to hypothesize a sequential establishment of a positive HAR over time: firstly with a decrease of attentive state and an increase of proximity and accepted contacts, and secondly with a

disinterest of human contacts and the expression of natural foraging behaviour. The latter may require a higher exposure time.

“We may be able to hypothesize a sequential establishment of a positive HAR over time: firstly with a decrease of attentive state and an increase in proximity and accepted contacts, and secondly with a disinterest of human contacts and the expression of natural foraging behaviour. The latter may require a higher exposure time.” Line 573

Line 514: Double reference

Yes, corrected.

Line 568: More likely to? But does this not show that this test or variable is not generalisable?

Not entirely, even if the identity of the human was in the model, the statistical analysis still found effects of the treatment / proximity / time. But it is true that we need more information to understand what are the causes of the effect of the human. We had a sentence on this point in the previous version (kept in the second).

“Thus, more studies of human features that are most likely to generate a positive HAR are needed and may be of interest regarding animal welfare. In addition, studying human-piglet relationship in a more systematic way, as in other domestic species, for example the play behaviour in dogs (Horowitz & Hecht, 2016) or the pet directed speech (Jeannin et al., 2017; Lansade et al., 2021), may shed light on the evolution and converging strategies of interspecific relationships. However, the influence of human identity did not modify the general outcomes of our study, but only decreased some effects, suggesting that this variability does not modify the main results, but should be considered in future studies” Line 705

Line 577: The major difference to the pigs is more likely to be in the difference in smells of the two handlers. Was there any thought given to soaps and perfume? Even body odours differ.

See response above.

Lines 586-588: But some variables were not significant for one of the handlers.

It depends on the type of test and the variable. Only the first behavioural score ReuPC1 of the Isolation/Reunion showed a significant interaction between the humanID and the Treatment and it was not found at all for the sessions of additional positive contacts. Some vocal scores changed depending on the human but not in interaction with the treatment. We report these findings and discuss them to encourage the community to run more controlled experiments to test what is making different humans perceived differently, we are not claiming we have demonstrated any causes. As we said earlier, the fact that the identity of the human was in the full model means that this variability was in the model, and do not rule out other significant effects. We added more information in the result section

- “During the reunions of the Isolation/Reunion test, the interaction between treatment and human identity was significant for the first behavioural proximity score (ReuPC1,  $\chi^2_1 = 6.01$ ,  $p = 0.01$ ) but not the others (ReuPC2 and ReuPC3 ( $\chi^2_1 < 1.98$ ,  $p > 0.16$ , table S1).” Line 496

- “These interacting effects of the human identity and treatment on behaviour were not found when considering the reunions of the conditioning ( $\chi^2_1 < 1.32$ ,  $p > 0.25$  for all CondPCs, table S1).” Line 501

- “Interactions between the human identity and conditioning time were not significant, neither considering the reunions of the Isolation/Reunion test (ReuPCs,  $\chi^2_1 < 0.642$ ,  $p > 0.42$ , tables S1), neither the trial number during the session of additional positive contacts of the conditioning (CondPCs,  $\chi^2_1 < 0.11$   $p > 0.74$ , table S1).” Line 504

And in the discussion:

- "The effect of the human did not interact with the conditioning time, leading to the conclusion that the difference between the two experimenter may have establishment during the period of positive handling at weaning, prior to the conditioning." Line 684

Line 600: "We suggest that the use of vocalisations to assess quality of human-pig relationship could help to better monitor the parameters involved..." I don't know what this means?

The sentence was changed. See line 725

"We suggest that analysing vocalisations structure may be a good tool to assess the quality of human-pig relationship and help monitor the establishment of a positive HAR."