



University of
Zurich^{UZH}

Institute of Evolutionary Biology
and Environmental Studies

University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich
Phone: +41 44 635 52 71
Fax: +41 44 635 54 90
www.ieu.uzh.ch

UZH, IEU / Animal Behaviour, Winterthurerstrasse 190, CH-8057 Zurich

David A.W.A.M. Jansen
PhD student
Phone: +41 44 635 52 81
david.jansen@ieu.uzh.ch

Zurich, October 26, 2023

MS: 3399353257711615 - Revised for BMC Biology

Temporal segregation of individual and behavioral signatures in banded in banded mongoose close calls

David A.W.A.M. Jansen, Michael A. Cant and Marta B.

Dear Kester Jarvis,

We would like to submit our revised manuscript "Temporal segregation of individual and behavioral signatures in banded mongoose (*Mungos mungo*) close calls" by David A.W.A.M. Jansen, Michael A. Cant, and Marta B. Manser for publication as research paper in BMC Biology.

Our thanks go to the reviewers and the editor for their thoughtful critiques of our manuscript. We have adopted most of the suggestions, including clarifying the term vocal signature versus vocal cue, the uniqueness of our findings, and incorporated the within-syllable aspects as suggested by the 5th reviewer. Based on the comments and our changes we have also edited the title of the manuscript. The new title of the paper is:

"Segmental concatenation of individual signatures and context cues in banded mongoose (*Mungos mungo*) close calls"

We think that the manuscript has greatly improved by these revisions and we hope that you will find it suitable for publication in BMC Biology. Our point-by-point responses to comments are outlined in detail on the following pages, and we also refer to the changes in the manuscript.

Sincerely yours,

David A.W.A.M. Jansen*, Michael A. Cant, and Marta B. Manser

*Corresponding author: David A.W.A.M. Jansen; Animal behaviour, Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.
Phone: +41446355281, fax: +41446355490, email: david.awam.jansen@gmail.com



Queries/critiques are numbered and in red, italic font.

Responses follow in black font. We have also indicated on which page (P) and/or lines (L) changes have been made.

Revised prose from within the manuscript is in blue font.

General response

#G1. Several of the authors point out that the term 'signature' is used in a non-standard way and reviewer 5 points out that vocal cue would be more appropriate when referring to non identity based signature. We have changed and clarified this in the abstract (P2), added vocal cue as keyword (P2), and first paragraph of introduction (P3, L5-10) and refer to vocal signature and other vocal cues as 'vocal cues' in the rest of document.

#G2. We realized that the mixed use of 'element', 'part' and 'segment' could be confusing. We have now consistently used the term 'segment' in the manuscript.

Detailed responses to Reviewer #1

Background

#1.1 'Vocal signatures' needs to be defined. Do you mean 'calls with individually distinctive acoustic properties'? If so, why not say this and then say (for simplicity, 'vocal signatures').

We have included this now (P3, L5-10) in addition to the changes explained in #G1.

#1.2a Methodology/Principle findings

'The first evidence'? This is an exaggeration. What about the grunts given by vervet monkeys that differ depending on whether the animal is approaching a dominant, a subordinate, moving into an open area, or seeing another group? The issue of variation along two dimensions addressed in this study by first showing that one individual varied specific acoustic features from one context to the next and then showing that, despite inter-individual differences, other individuals varied the same acoustic measures. And what about the grunts of baboons (Owren's papers) that are both individually distinctive and acoustically different depending upon whether the vocalizer is on a move or approaching a female with infant? There are many such examples of acoustic subtypes. This section needs to be rewritten.

We acknowledge that multiple previous studies (including the meerkat work, and we have added the baboon study as such an example, P3, L19-21) have shown the presence of multiple signatures/cues in vocalizations. In our study, we do not argue that the presence of multiple vocal cues related to different traits or context is a new, or unique finding. However, the previous studies imply that multiple cues are encoded in the frequency related structures of the whole call, and not, like in the close call of banded mongoose, in temporally separated segments. We agree that many primate calls are both individually distinctive and context specific, but (as also pointed out by reviewer 4) the identity 'signature' in these species' calls are not a temporally separate segment. Rather, the frequency related acoustic parameters over the whole call changes if, for instance the context changes. Overlapping signatures/cues potentially affect the clarity and reliability of the cues as pointed out by Briefer et al 2010. The work presented here and previous work on some bird species show that some cues can also be separated in time (temporal separation). We argue that these are examples of segregation of information as suggested by Marler in 1960. A similar process seems to occur in for instance the túngara frog and potentially several other non-linear calls. So in summary we do not argue that we provide first evidence for multiple signals, but for the first quantification of temporally separated cues in a single call. As pointed out by the other reviewers the term call versus for instance song is vague and therefore (as suggested by reviewer 5) we have emphasized the single syllable characteristic of the banded mongoose close call. The banded mongoose is the first described example of segmental concatenation encoding multiple cues within a single syllable call. Please see responses to reviewers 3 and 5 for specific adaptations.

#1.3 Conclusions/Significance

What do you mean by 'temporal segregation'? Calls given at different times? In different contexts? Once again, this is by no means 'the first evidence' of its kind. This is extremely over-stated! 'We argue that' but this idea has been in the literature for decades, ever since the call subtypes in vervet monkey grunts were first shown to exist through playback experiments. Of course you can argue that call subtypes allow a signaler with a limited vocal repertoire to overcome such limits and send many messages 'this is obvious' but the important question is: why so few subtype?' Why not more?

By temporal segregation we mean that within a single syllable call the acoustic structure differs in two distinct concatenated segments, of which the first segment correlates with individuality, while the second segment with the behaviour of the caller.

We have additionally emphasized that the banded mongoose close call is the first non-human example of within-syllable encoding. (P

#1.4 'Animals can encode'? It's not at all clear what you are claiming here, in large part because acoustic signatures is not defined. Is your point that acoustic signatures are rare? If so, this is simply wrong: they've been known for a long time, particularly in primates. Is it that acoustic signatures based on temporal features are rare?' If so, this is also incorrect because variation in temporal features help define the subtypes of vervet grunts, and probably the context-specific subtypes of many other species.

We hope to have clarified our definition of vocal signatures (and cues). See G1. We neither claim that acoustic signatures/cues or acoustic signatures based on temporal features are rare. We acknowledge that all of these findings have been previously reported (see for instance P3 L8-16, P3 L19-21 & P 4 L9-11)

#1.5 'However because acoustic space is limited and many acoustic parameters are correlated with one another, the amount of variation that can be used by signalers to encode different signature types is ultimately constrained.' Why? It's not in human speech, where speakers use a very narrow bandwidth and acoustic parameters are highly correlated.

To our knowledge human speech is also limited in variation of frequency related aspects (e.g. we only have a limited number of vowels and stop consonant). It's the phonological and syntax combinations that make human language virtually unlimited.

#1.5 'there currently exist no studies that refer to it within a single call type.'

This, again, is simply wrong, both in the empirical details and in the claim that the 'trade-off' argument has never been proposed before. In fact, Rendall & Owren have worked for years on the problem of how individual signalers simultaneously encode caller identity and context-specific variation in their calls. This is what has led them to argue that, in primates at least, individual identity is encoded primarily in formant structure whereas context-specific information is encoded in other acoustic parameters. They support this argument with extensive data on baboon grunts.

The studies by Rendall & Owren are very convincing examples on how multiple cues are expressed within the same call in different frequency related acoustic parameters. We have now brought in a reference to them (P3, L19-21). However, our study is about the quantification of multiple cues being expressed in temporally separated acoustically distinct call segments of a single syllable vocalization. We hope, we have clarified this aspect now much better (P4, L3-5 & P6 L14-22)

#1.6 Further, when Green presented his original data on Japanese macaque coos, he was criticized in a subsequent paper by Lillehei & Snowdon, who emphasized the importance of controlling for individual variation and looking simultaneously at the encoding of caller ID and context.

We did control for individual variation in while testing for the behavioral signature. By using a pDFA we had a test factor (in this case behaviour) and a control factor (individual), which has to be controlled for since, we had multiple calls per individual (in multiple behavioral classes). We have clarified our statistical analysis approach by rewriting part of the methods (specifically P10, L18-20 & P11, L9-11) .

#1.7 Finally, it's not at all clear why you are making this claim without mentioning Manser's work on meerkats alarm calls, where two types of information are simultaneously encoded in a single vocalization.

We do not think that the Manser's work on meerkat alarm calls is applicable to the work we present here. The meerkat alarm calls encode multiple types of information in frequency related acoustic parameters over the whole call as the grunts in vervets and baboons, and not in temporally separated segments we describe here for the close call of the banded mongoose.

#1.8 The idea that the data presented here are the 'first of their kind' is simply wrong. It greatly distracts from the paper.

We hope that with the changes of the scope and wording it becomes clear that the close call of the banded mongoose and how the multiple signatures are encoded does provide some unique findings. Moreover, our study provides the first example of a discrete individual 'segment' in a graded call containing information regarding individuality." See responses to #1.2 & reviewer 5 for details.

Detailed responses to Reviewer #2

#2.1 The authors emphasize the paper's novelty by repeatedly stating that it is the first example of an animal encoding multiple signatures into different elements of a single type of signal. Yet, they provide examples of two other well-studied systems that also encode multiple signatures into a single signal. White-crowned sparrows encode individual identity and group membership into temporally separated elements of a single song (this paper specifically discusses Marler's segregation of information hypothesis), whereas túngara frogs encode species identity and individual quality into contiguous, and often partially overlapping, elements of a single advertisement call. Another example is provided by meadow pipits, which encode individual identity into the beginning of their song and species identity into the end of their song (Elfström 1990 Anim Behav 40:786-788). The authors should state clearly how their findings are different from those of these other studies. Is it because this is the first example in a mammal, because the elements containing the signatures are contiguous (i.e., not temporally separated, as they are in bird song), or because the other examples aren't considered a 'single call type'? The whine and chuck series of the túngara frog's advertisement call would certainly qualify as a single signal, and most ornithologists would also consider the song of the white-crowned sparrow and the meadow pipit to be a single signal, despite having multiple temporally separated elements, such as trills and note complexes. This does not diminish the value of the current study, but the differences between it and these other studies should be stated more clearly.

As pointed out by reviewer 5 the uniqueness of the banded mongoose close call is the within-syllable encoding. We acknowledge that this was not clear in the original manuscript and we changed the manuscript to clarify this. (see responses to reviewer 5 for details).

#2.2 I applaud the authors for using a permuted discriminant function analysis (pDFA) that properly accounts for non-independence among multiple calls from the same individuals... This is very rare in animal behaviour research, and is one of the greatest strengths of the current paper. Although I have no concerns about the general statistical approach, I do have some suggestions to help clarify the description of the statistical methods....

#2.2a First, the purpose of the linear mixed model is a bit unclear. Was it performed solely for the purpose of calculating variance inflation factors and obtaining a subset of acoustic parameters that was free from multicollinearity for use in the pDFA, or was it used in addition to the pDFA to compare calls among behavioral contexts? Were the acoustic variables used as covariates in a single linear mixed model (if so, what was the dependent variable), or were they used as dependent variables in separate models?

We used the linear mixed effect models predominately the purpose of calculating variance inflation factors and obtaining a subset of acoustic parameters that was free from multicollinearity as this is essential for the proper functioning of the pDFA. Additionally we used linear mixed effect models to test a small subset of the variables (e.g. duration) to compare calls between behavioral contexts. Variables were included as covariates and both group and individual identity (nested in group) were included as random factors. We have changed this part in the methods to clarify this (P10, L9-14).

#2.2b Second, it is unclear exactly how many pDFAs were conducted. I understand that the whole call, the noisy base of the call, and the harmonic part of the call were analyzed separately. But, for each of these call components, did the authors conduct a single pDFA that included multiple factors for individual, behavioural context, group, and sex (i.e. a total of 3 pDFAs), or were separate pDFAs conducted for each group and sex? The fact that Table 1 presents separate classifications for each group suggests that separate pDFAs were conducted for each group and for each sex (i.e. 3 call components x 2 sexes x 4 groups = 24 pDFAs). If this is the case, do the results presented in the results section correspond to males or to females. One or two sentences near the end of the statistical analysis section would really help to clarify how many analyses were used, and which factors were used in each analysis.

We performed 4 pDFA to test for overall and the pairwise comparison between behavioral context. In addition we performed 2 additional pDFA to test for the group signature (while controlling for individual) and for sex cue (again while controlling for individual). Therefore we in total performed 6 pDFA. As suggested we have added this in two additional sentences in the statistical analysis paragraph. (P11, L7-11).

#2.3 At present, there is no way to assess which acoustic variables were included in the pDFAs following the test for multicollinearity and the stepwise selection procedure, or which of the variables in the final models had the greatest effect on the discriminability of individuals and behavioural contexts. If only 3 pDFAs were conducted (see comment 2 above), then the authors could include a small table (perhaps as an 'Additional Material Files') that shows the final discriminant functions and their corresponding loadings. If more pDFAs were conducted, then perhaps the authors could include a sentence or two that states, in general terms, which variables were used in the final analyses and which were the most important for discrimination. This is also important because an assumption of the pDFA is that the number of variables included must be smaller than the number of individuals included in the smallest class. Groups 11 and 15 had only 7 individuals, so the corresponding analyses should have included no more than 6 predictor variables.

The exact number of variables (and the types) that we included in the various analyses was different per test. We are aware that the number of variables included in the pDFA should be smaller than the individuals and this was ensured in our analysis. We have added some sentences to clarify this in the manuscript. (P10-11 L30 -2)

4. The authors should mention that future research should use playback experiments to test if the behavioural context signature is important to receivers. This could go in the first paragraph of the discussion following the sentence that describes receiver responses to individual signatures.

We have added a sentence to emphasise the importance of a playback experiment. (P6 L12-13)

2.5a. The acoustic analysis methods suggest that there are 19 acoustic variables (16 from the automatic measurement procedure, 2 derived variables, and the number of pulses), yet Table 3 only shows 14 variables.

In the final analysis we only used the 14 variables. We have changed the number in the text accordingly. The number of pulses was not used in the analysis and therefore we have removed the sentences about the pulse analysis. (P9 L21)

2.5b. Also, some of the acoustic variables are not described in sufficient detail to be replicated. 'Bandwidth', 'Maximum Frequency', and 'Minimum Frequency' need to be defined relative to the call's 'peak amplitude' (this value is available directly from Avisoft's automatic measurements dialogue box). For example, maximum frequency could be defined as the highest frequency that was within 10 dB of the peak amplitude. Otherwise, minimum frequency would always be 0 Hz and maximum frequency and bandwidth would always be the highest frequency of the recording (probably 22050 Hz).

Does 'mean' indicate that measures were obtained from an averaged spectrum, or that they were obtained from every spectrum within the call and then averaged. These two methods produce very different results, so the method should be defined clearly in the text. For example, the authors could state 'all frequency measures were obtained from the mean spectrum of each call or call component,

and the 3 quartiles were also measured from the point within the call or call component that had the maximum amplitude.'

Finally, the authors should explain where within the call the frequency measurements were obtained. Table 3 states 'mean' or 'max' in parentheses, but these terms are unclear and will not be meaningful to anyone who is unfamiliar with Avisoft software. Does 'mean' indicate that measures were obtained from an averaged spectrum, or that they were obtained from every spectrum within the call and then averaged. These two methods produce very different results, so the method should be defined clearly in the text. For example, the authors could state 'all frequency measures were obtained from the mean spectrum of each call or call component, and the 3 quartiles were also measured from the point within the call or call component that had the maximum amplitude.'

We have added some sentences to clarify how and where acoustic variables were measured. (P9-10 L22-1).

#2.5c. Also, the frequency variables listed in Table 3 should have 'Hz' listed in parentheses as their unit of measurement.

We have added Hz to all frequency related measurements in table 4. (Was table 3, see #2.6) (P21)

#2.6. Table 4 should be re-labelled as 'Table 1' since it is mentioned first in the text.

We have re-labeled Table 4 (overview of behaviors) to table 1 and accordingly adapted numbering of other Tables and references to tables in text.

Table 1 (overview individual classification)	→ Table 2 (P19)
Table 2 (overview behaviour classification)	→ Table 3 (P20)
Table 3 (overview acoustic variables)	→ Table 4 (P21)
Table 4 (overview behaviour)	→ Table 1 (P18)

#2.7. The footnotes in Tables 1 and 2 state that results are included for 'the whole call and the noisy part' of the call, but should also state that results are included for the 'harmonic part' of the call.

We changed the footnotes and added the harmonic part. (P19 & P20)

#2.8 In the recording methods, please state the file format, number of bits and sampling rate used in two recorders. We added a sentence to clarify these details. (P9 L8)

9. I have a few suggested wording changes and typo corrections.

In the last line of the Background in the Abstract, 'emitted close graded close calls' should be 'emitted graded close calls.' In the last sentence of the Introduction, delete 'in their close calls' and change 'aspects of a graded call' to 'aspects of this graded call.' In the results, change 'considerable' to 'considerably' and change 'neither' and 'nor' to 'either' and 'or' (or change 'No evidence' to 'Evidence'). In the Discussion, 'tungara' should have an accent above the 'ü'. In the fourth paragraph of the Discussion, change 'Similar' to 'Similarly' and change 'unambiguity' to 'clarity'. In the last sentence of the Discussion, change 'avoid the lack of ambiguity' to 'avoid the ambiguity.' In the Conclusion, change 'provide equally' to 'provide an equally' and change 'in animals, and in humans' to 'in human and non-human animals.' In the first line of the Recording Methods, should it be 'more than' instead of 'less than' In reference 48, delete the second 'd' from 'handbookd'. In Table 1 footnote, change 'change' to 'chance.' Throughout the Methods, change the 3 instances of 'analyzes' to 'analyses' since this in the noun form of the word.

We have made all the suggested changes.

Detailed responses to Reviewer #3

General comment. The term 'signature' is typically used to describe information encoded in vocalisations about the identity of an individual (e.g. species, group, individual), similarly as in humans ('A distinctive mark, characteristic, or sound indicating identity'). It does not sound right to me to use it in order to describe behavioural context, particularly because the context in which an individual calls is not fixed over time. Therefore, I suggest changing 'signature' to 'information' or 'cues', or else, when referring to behaviour/context throughout the ms. Note that the literature cited in the background P3, L5-7 did not use the term 'signature' to describe vocal indicators of male quality or reproductive state.

We have addressed this point (see #G1).

#3.1 P3-L21. I would emphasize here that for individual recognition, we need variation between individuals (and stereotypy within individual), not just 'variation'.

We have changed the sentence to emphasize the need of between category variations. As the same principle counts for other cues and signature information (e.g. group, behaviour or sex cues) we kept the sentence general instead of referring to the individuality (P3 L16-18).

#3.2 P3-L24. I would mention here that segregation of information can be done by using either different features (temporal segregation) or different ranges of the same feature (e.g. frequencies) to convey different information (Marler 1960).

We have changed the sentence to clarify that segregation of information is not necessarily temporal, but can also be spectral (e.g. using different frequencies. (P. 3 L28-30)

#3.3 P3-L26. Another good bird example of temporal segregation is the Ortolan bunting, in which the initial song part codes for the individual identity and varies between males and the final part codes for the population identity (e.g. Osiejuk et al. 2005, Behavioural Processes 68:69'83).

We acknowledge that this study also encodes multiple signatures, but we have already mentioned several other studies on this aspect.

#3.4 P5. I suppose that all the ncce provided in the results are percentages, not just numbers' If yes, '%' could be indicated after each of them.

The ncce values provided are not exactly the same as percentages and can therefore not be directly compared with for instance the correctly cross validated percentage from more traditional DFA's. The ncce values are the results of multiple random selections and permutations.

#3.5 P5-L8. This ncce must be wrong, it is exactly the same as the one mention above and the p value is significant (contrary as what is written in the text).

The values are correct as are the p values. Please note that the significance of the values is based on the number of classes and individual presents in those classes. Therefore the value at which it is significant will differ.

#3.6 P5, L11-19. It would be useful if contexts could be described before these results, in the introduction, or at least, mention Table 4 here.

We have added a reference to the table in results session. (P 5 L17)

#3.7 P5, L16-18. Table 2 only shows general differences between contexts. If such claims are being made here (e.g. duration increases, etc.), then statistical results comparing each acoustic parameter value between contexts should be provided. Such tests are mentioned in the methods (linear mixed effect models), but do not seem to appear anywhere in the results.

The linear effects models were used to calculate the VIF values see also #2.2. The differences between the duration are given in the results session. (P5 L20-23)

#3.8 P6, L1-3. This is the case if individuals from a given group are more genetically related than individuals from different groups. The other mechanism that can arise when individuals change groups and in species capable of vocal flexibility is group convergence (e.g. Knörnschild et al. In Press. Animal Behaviour; Briefer and McElligott 2012, Animal Behaviour 83, 991-1000).

We acknowledge that this is a possibility and have added it to the text. Currently there is not data available for the banded mongoose to investigate which of these two processes would be applicable. (P6 L9-11)

#3.9 P6, L5-13. This kind of segregation of information is not rare in songs of birds or even mammals. The difference between 'songs' and 'calls' is quite subjective (at least 'human subjective'). The term 'song' is usually used for vocalisations composed of sequences of elements. This presence of multiple elements facilitates temporal segregation of information. 'Calls' is used for simpler vocalisation composed generally of one single element. If only one simple element is present, temporal segregation of information is not possible. These definitions of 'songs' and 'calls' established by researchers are the only reason why this paper is the first to find segregation of information in a call. See for example hyrax, which produce a complex call termed 'song' by researchers. This song encodes information regarding individual identity, body weight, size and condition, social status and hormonal state in different components (Koren and Greffen 2009, Behav Ecol Sociobiol 63:581'590). Therefore, I suggest revising this claim on the novelty of the paper. This would not make the paper less interesting than it is, because it still shows a very nice case of segregation of information.

We have explained in more detail this study is about the temporal segregation of vocal cues in acoustically distinct segments within the same syllable of the call. (see #1.2 and responses to reviewer 5 for details).

#3.10 P8-L14. What do you mean by 'less than 1 year'?

This was an error and should have been "more than". We have changed this. (P9 L4)

#3.11 P9, L2. Maybe mention what is the lowest fundamental frequency measured in banded mongoose here, to justify the filter.

For final analysis the filter was not used. We have removed the sentence about the filter.

#3.12 P9-Acoustic analysis. Some parameters like Max. Freq. and Min. Freq., listed in Table 3 are not described in the methods.

See #2.5

#3.13a P9. Statistical analyses. I did not find where the results of the linear mixed effect model were. It would be useful to present them in a table (e.g. in Table 3).

Linear effect models were not used to determine differences in acoustic variables, but to determine VIF values. We have adapted the text to clarify this. See #2.2a and #2.3

The part on parameters with VIF '2.5 included in the analyses or not is not clear. In which analyses where these parameters included' 'the remaining parameters were entered into a DFA': does that mean that the parameters entered in the DFA were multicollinear? Can these parameters included in the DFA be listed somewhere? One of the assumptions of DFA is an absence of perfect multicollinearity, which means that parameters entered in the DFA should not be too redundant. If this is the case, a PCA can be carried out before the DFA. Is that the case here?

We did not perform a PCA. Only parameters with VIF values below 2.5 were included in the analysis to avoid problems with multicollinearity. See also See #2.2a and #2.3. We have adapted the part on the VIF analysis and variable selection slightly to clarify the analysis.



How were correct classification probabilities determined for behavioural context while controlling for individual and vice-versa? Was a permuted DFA used (pDFA is mentioned in the methods, but later on)? More generally, it is not very clear how sex, group, behavioural context and individuality were all controlled for in the pDFA, given that a pDFA can control for only one factor at a time (one test and one control factor). From the results, it seems that groups were analysed separately, which seems ok, but then this should be explained in details in the methods.

For the analysis of the vocal cues related to behaviour, group and sex we controlled for individual. Group and sex cue analysis were not controlled for behaviour. We have moved the explanation of the pDFA forward. P10 L10-14).

#3.14 P10,L3-4. Which criteria on the basis of results from the linear effect models were used to select parameters?

We used the mixed effect models to calculate VIF values. Only VIF values of below 2.5 were included in the analysis. (P10 L9-11 & L15-18; See also #2.2, #2,3 and #3.13)

Detailed responses to Reviewer #4

REFEREE 4

Reviewer's report

I agree the novelty of the findings seems to rest on the definition of a call vs a song and the importance of the temporal segregation of the info (as reviewer 1 points out many primate calls are context specific and individually distinctive, but the identity 'signature' in these calls is not a separate temporal element, so personally I don't think the vervet findings are directly comparable). I think the results need more firmly embedded in the existing literature (where it is more explicit as to why these findings are different or similar to other findings), however I do think this will still be a paper of wide interest to biological researchers. Complexity of communication systems seems to be a topic attracting wide interest across taxa currently, so I think these findings remain interesting and suitable for BMC Biology.

We think that by addressing the suggested changes and comments suggested by the other reviewers we have addressed the above comments. See specifically #5.4

Detailed responses to Reviewer #5

REFEREE 5

Reviewer's report

This interesting paper demonstrates that different types of information are encoded within different temporally-segregated portions of the same short contact call, in banded mongooses. Specifically, the initial noisy part of the call encodes individual and group identity, while the (optional) more harmonic second part encodes the behavioral context, or current activity of the animal.

I think this paper is interesting and worth publishing, but I agree with the first reviewer that its overarching presentation and contextualization need some work. Specifically, the non-specialist reader will be left wondering what, exactly, is new about this finding and why it is important.

#5.1 There are two problems: the first is the use of the term "signature" which is used in a non-standard and, particularly in this system, non-helpful way. "Signature" is typically used to discuss acoustic cues to individual and/or group identity, and NOT sex, size, quality etc. For the latter "vocal cues to" or "acoustic indicators of" would be more standard or more appropriate in this case.

See G1.

#5.2 The second is that "temporal segregation" is too vague: this would include sequential combination of separate acoustic entities of any size or continuity. Thus, many complex birdsongs encode individuality in the song, specifically in the sequence of individual song syllables (e.g. nightingales), or the co- and qui- notes of the coqui frog encode male- versus female-directed information. Such encoding has also been shown in cetaceans and primates and is nothing new, even in mammals.

#5.3 What IS new is that a single short syllable contains two different types of information, but in temporally-segmented fashion. This is the way human consonant-vowel syllables ("ba" vs "pa" /"ba" vs. "bo") work, and this seems a far more appropriate analogy than the syntactic one ("I John eat" versus "I John move") used here. So the key novelty is in the within-syllable encoding aspect of the findings presented here, which might be termed segmental concatenation, or "phonological". It is the within-syllable aspect of these findings that is novel and important.

#5.4 So my suggestion is to reframe the result in these terms:

"Segmental concatenation of individual signatures and context cues within single syllables of banded mongoose contact calls"

"While many animal signaling systems use concatenation of acoustically-separate syllables to enrich and extend the signaling space (e.g. birdsong), human speech also uses a different type of encoding into individual syllables, at a phonological level. Thus a stop consonant like /b/ versus /p/ can be combined with a vowel like /a/ or /o/ to create a richer signalling unit than either class alone could provide. Such combinations (versus "syntactic" concatenation of syllables and words) are a core component of the phonological component of human spoken language"

We agree that the single syllable approach is a good one and clearly emphasizes the novelty of this study. We have emphasized the fact that the close call is a single syllable call throughout the paper and changed the title, abstract and discussion accordingly.

New title: [Segmental concatenation of individual signatures and context cues in banded mongoose close call](#)

We have additionally slightly restructured the discussion. P6 L14-23

The authors might also cite that specific segment types commonly signal group (dialect) differences in human speech (e.g. trilled versus uvular /r/ in German dialects, vowel identity in English dialects) which is close to their "group signature" idea.

Of course, the analogy is imperfect since humans don't (typically) have individual specific segments (though an individual lisp or similar abnormalities might reliably signal identity in some cases). But nor do humans walk around saying "I John eat". No analogy is perfect.

We acknowledge that this is a nice analogy, but as we think the group signature is not the exciting part of the paper we have not included it in the manuscript. We think it would add unneeded complexity to the manuscript.

While this way of framing the results doesn't do full justice to either human syntax or to phonology, it at least will make clear to most readers what this study is all about.

I think the authors should present spectrograms as Figure 1, with some rather extreme examples to illustrate the phenomenon.

The figure with the spectrograms of the calls related to the different behavioral context and the consistent individual related first segment seems to us to illustrate the phenomenon within the banded mongoose calls. We could bring in a figure with the mentioned examples of species that show the same phenomenon, e.g., tungara frogs, etc...however, we need to ask for permission there first. Currently, we just refer to the publications, where this spectrograms are available.