**Reviewer #3**

**Questions**

* **1. Summary Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).**
  + This research delves into understanding Shiba Inu dog vocalizations using a unique dataset sourced from YouTube. Differing from traditional methods, this study employs a data-driven approach by extracting dog sounds and their context from videos to identify patterns. The findings reinforce earlier research, such as growls indicating interactions, and introduce new insights. For example, a "whimper" may signify attention-seeking or discomfort, and distinct meanings were identified between "bow-wow" (indicating curiosity) and "bark". The team developed a versatile framework to process dog videos from YouTube, suitable for other animals too. This innovative approach offers a blend of existing knowledge and new insights into dog vocal semantics.
* **2. Strengths and Weaknesses Please provide a thorough assessment of the strengths and weaknesses of the paper, touching on each of the following dimensions: - Novelty (how novel are the concepts, problems addressed, or methods introduced in the paper) - Quality (is the paper technically sound?) - Clarity (is the paper well-organized and clearly written?) - Significance (comment on the likely impact of the paper on the AI research community, i.e., explain how the paper might impact its own sub-field or the general AI community.).**
  + Strengths and Weaknesses Assessment:  
      
    1. Novelty:  
    - Strengths: The paper introduces a unique data-driven approach to deciphering dog vocalizations, moving away from the traditional method of directly recording dog sounds. The utilization of a dataset sourced from YouTube to extract dog vocalizations and their context is innovative. The study's distinction between previously conflated sounds (e.g., "bow-wow" vs "bark") showcases an attempt at a more nuanced understanding.  
      
    - Weaknesses:  
    The core idea of understanding animal semantics isn't new, though the methodology is.  
      
    2. Quality:  
    - Strengths: The research is grounded in a well-constructed pipeline that extracts and analyzes dog-related videos. There's an emphasis on contextual information, with the identification of up to 11 locations and 14 activities.  
      
    - Weaknesses: Sole reliance on YouTube might introduce biases or miss out on certain nuances present in controlled experimental settings. The generalizability of the findings to other dog breeds or animals isn't thoroughly discussed.  
      
      
    3. Clarity:  
    - Strengths: The paper appears to be well-organized, detailing the methodology and the significance of its findings. Key contributions are succinctly summarized, aiding understanding.  
      
    - Weaknesses: More elaboration on the challenges of extracting meaningful sounds and context from videos would be beneficial. The mention of "Figure 1" without further explanation in the provided text can be confusing for readers. Moreover, the authors utilize Sentence Split, Word Segmentation, and Subword Extraction based on the assumption described. It would be beneficial for the readers if empirical evidence or precedents supporting this assumption were provided.  
      
    4. Significance:  
    - Strengths: Given the first-of-its-kind approach, this paper could inspire a new wave of animal language studies leveraging social media and web data. The universal pipeline proposed is reusable, making it valuable for researchers aiming to study other animals.  
      
    - Weaknesses: Its significance might be limited to those interested in the semantics of animal vocalizations. The broader AI community may not find direct applications, though the methodology could be adapted to other domains.
* **3. Questions for the Authors Please carefully list the questions that you would like the authors to answer during the author feedback period. Think of the things where a response from the author may change your opinion, clarify a confusion or address a limitation. Please number your questions.**
  + 1. Data Construction:  
    - For data collection, how have the vocalizations from multiple dogs been managed or integrated?  
    - The authors proceed with Sentence Split, Word Segmentation, and Subword Extraction based on the assumption described below. However, this assumption seems quite strong. Are there any empirical evidences or precedents that support it?  
    "Similar to human language, where words serve as the fundamental units for constructing sentences, we hypothesize that a similar concept can be applied to dog language. We define a “word” as an independent and contiguous dog vocal sound that typically lasts around 1 second, and it is bounded by some noticeable pauses. A “sentence” is a sequence of consecutive words. A “subword” is a subpart of “word” and is represented by an IPA symbol."  
    - How did you ensure the accuracy and consistency of defining words, sentences, and subwords in dog vocalizations?  
      
    2. Methodology:  
    - Considering the diverse range of dog breeds, why was Shiba Inu chosen as the primary focus?  
    - How does the framework account for the potential influence of human interaction or commands when analyzing the dog vocalizations from the videos?  
      
    3. Results:  
    - The same Word Type (e.g., Bark) possesses different Phonetic Symbols. What implications does this variation hold?  
    - How do you account for potential variability or deviations in dog vocalizations based on breed, age, or individual personality?  
    - With the understanding that specific Word Types can have different Phonetic Symbols, how do you envision this influencing real-world applications, such as training or communication with dogs?  
    - Were there any unexpected findings or vocal patterns that did not align with the initial assumptions or hypotheses?
* **4. Reproducibility Does the paper provide enough information to be reproducible? If not, please explain (It may help to consult the paper’s reproducibility checklist.)**
  + Yes
* **5. Resources Are there novel resources (e.g., datasets) the paper contributes? (It might help to consult the paper’s reproducibility checklist)**
  + Yes
* **6. Ethical considerations Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias, etc.)? If not, explain why not. Does it need further specialized ethics review?**
  + yes
* **7. OVERALL EVALUATION Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper.**
  + Weak accept: Technically solid paper where reasons to accept, e.g., good novelty, outweigh reasons to reject, e.g., fair quality.
* **8. CONFIDENCE How confident are you in your evaluation?**
  + Quite confident. I tried to check the important points carefully. It is unlikely, though conceivable, that I missed some aspects that could otherwise have impacted my evaluation.

**Reviewer #4**

**Questions**

* **1. Summary Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).**
  + The authors describe their work on analysing dog vocalisations in the way of determining the meaning of a number of dog sounds. A pipeline that automatically extracts dog sound events from a video file, recognises the location and type of activity (using object recognition) and segmentation of the sound, was developed. Correlations between "words" and meanings were then inferred using conditional probability.
* **2. Strengths and Weaknesses Please provide a thorough assessment of the strengths and weaknesses of the paper, touching on each of the following dimensions: - Novelty (how novel are the concepts, problems addressed, or methods introduced in the paper) - Quality (is the paper technically sound?) - Clarity (is the paper well-organized and clearly written?) - Significance (comment on the likely impact of the paper on the AI research community, i.e., explain how the paper might impact its own sub-field or the general AI community.).**
  + Strengths:  
    - The problem of understanding the meaning of dog sounds by analysing videos is underexplored thus the work proposed presents a novel contribution  
    - The paper is generally well-written although it could benefit form some proofreading  
      
    Weaknesses:  
    - The analysis is based on correlations measured using conditional probabilities; thus the core AI work (the different components of the pipeline) are based on pre-existing models/methods.  
    - The identification of some variables (action, location) for data points is based on automatic methods whose accuracy are not that high (e.g., 0.59 for action recognition).  
    - The significance and impact are quite limited to researchers working on animal sounds.
* **3. Questions for the Authors Please carefully list the questions that you would like the authors to answer during the author feedback period. Think of the things where a response from the author may change your opinion, clarify a confusion or address a limitation. Please number your questions.**
  + In Section 3.2, perhaps you can clarify what "sonority fluctuation" is? Otherwise it is not clear what you consider to be a subword.  
      
    In Section 4.2, you report the accuracy for the top 5 predictions of your activity classification model. Did you actually use the top 5 predictions in labelling your data, or would you select only the topmost prediction (in which case the labelling will only be 0.59 accurate)?
* **4. Reproducibility Does the paper provide enough information to be reproducible? If not, please explain (It may help to consult the paper’s reproducibility checklist.)**
  + Yes
* **5. Resources Are there novel resources (e.g., datasets) the paper contributes? (It might help to consult the paper’s reproducibility checklist)**
  + No
* **6. Ethical considerations Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias, etc.)? If not, explain why not. Does it need further specialized ethics review?**
  + NA
* **7. OVERALL EVALUATION Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper.**
  + Weak Reject: Technically solid paper where reasons to reject, e.g., poor novelty, outweigh reasons to accept, e.g. good quality.
* **8. CONFIDENCE How confident are you in your evaluation?**
  + Somewhat confident, but there's a chance I missed some aspects. I did not carefully check some of the details, e.g., novelty, proof of a theorem, experimental design, or statistical validity of conclusions.

**Reviewer #6**

**Questions**

* **1. Summary Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).**
  + This paper presents a data-driven study of dog vocalizations from thousands of videos. The study uses correlations between words, subwords, locations, and activities to describe two aspects focusing on Shiba Inu dogs: 1) connecting vocal patterns with meanings, and 2) connecting vocal expressions with environmental factors that give rise to meanings. The results show strong correlations between certain words, between words and locations, and words and activities.
* **2. Strengths and Weaknesses Please provide a thorough assessment of the strengths and weaknesses of the paper, touching on each of the following dimensions: - Novelty (how novel are the concepts, problems addressed, or methods introduced in the paper) - Quality (is the paper technically sound?) - Clarity (is the paper well-organized and clearly written?) - Significance (comment on the likely impact of the paper on the AI research community, i.e., explain how the paper might impact its own sub-field or the general AI community.).**
  + Strengths:  
      
    S1) The paper seems to advance the state-of-the-art understanding of discovering semantic patterns in dog vocalization.  
      
    S2) The paper provides informative answers to the two posed questions.  
      
    S3) The proposed method of data collection, transformation, and analytics is sound and intuitive.  
      
    S4) The paper is relatively well-written and easy to follow.  
      
    Weaknesses:  
      
    W1) The data quality seems not convincing. Especially, the activity recognition by humans is 0.59, which is relatively low, and makes the findings that rely on this data questionable.  
      
    W2) The findings presented seem relatively trivial. The findings are claimed to be aligned with prior work, which is fine, but it is then unclear what is the novel insight brought by the present paper.  
      
    W3) The paper has certain inconsistencies and omissions in the writing. For instance, the steps in figure 2 are not directly aligned with the steps described in text. The classification task is pressumably single-label, but this is not obvious and needs to be discussed because, for example, the locations are not mutually exclusive: a dog could be in the living room with food nearby.
* **3. Questions for the Authors Please carefully list the questions that you would like the authors to answer during the author feedback period. Think of the things where a response from the author may change your opinion, clarify a confusion or address a limitation. Please number your questions.**
  + 1. Can you comment on the data quality? If I understand correctly, the activity classification by people is 59% - what does this mean for the analysis that relies on such an ambiguous data?  
      
    2. What are the main takeaways from the analysis? In other words, what does a researcher learn from this paper that they didn't know from other works?  
      
    3. Is the task single-label? If so, why?  
      
    4. Would the findings be expected to generalize to other dog species?
* **4. Reproducibility Does the paper provide enough information to be reproducible? If not, please explain (It may help to consult the paper’s reproducibility checklist.)**
  + Yes
* **5. Resources Are there novel resources (e.g., datasets) the paper contributes? (It might help to consult the paper’s reproducibility checklist)**
  + Yes
* **6. Ethical considerations Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias, etc.)? If not, explain why not. Does it need further specialized ethics review?**
  + There is a potential risk that the analysis is biased to the dog species of Shiba Inu.
* **7. OVERALL EVALUATION Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper.**
  + Weak accept: Technically solid paper where reasons to accept, e.g., good novelty, outweigh reasons to reject, e.g., fair quality.
* **8. CONFIDENCE How confident are you in your evaluation?**
  + Somewhat confident, but there's a chance I missed some aspects. I did not carefully check some of the details, e.g., novelty, proof of a theorem, experimental design, or statistical validity of conclusions.