Thanks for your careful and valuable comments. We are pleased to note that you have found our research work interesting and also pointed out some problems to help us improve the quality of our work. We will explain your concerns point by point.

**Common Concerns:**

**Q: The meaning of “barks” and the unvoiced parts of barks.**

A: Unvoiced sounds are made when the vocal cords are not vibrated. In this paper, we take “barks” in its broadest sense, which represents any vocal expressions coming from a dog, including bark, whine, howl, huff, growl, yelp, and yip. In this case, it’s possible for dogs to generate unvoiced parts in their vocal expressions as above. We sincerely apologize that we didn’t state the definition clearly in the paper to make you misunderstand.

**Q: The influence of other confronting factors like body size, age, sex, mood, recording condition.**

A: We agree that some factors including dogs’ body size, age, sex, mood, and recording condition will bring a difference to their barks. However, we have emphasis that these factors are not related to the cultures of different nations. As we have built our dataset from a wide range of sources, the distributions of these factors will not vary between two language environments. At the same time, noises ~~both~~ exist in both environments. We have tried our best to reduce impact by enlarging the population of dogs in our dataset. The influence of these factors is interesting for sure, we may take research on that in future work.

**Reviewer#1:**

**Q: The meaning of unvoiced parts of barks in human languages.**

A: The variable of “unvoiced segment length” is defined in the original paper ([url](https://ieeexplore.ieee.org/abstract/document/7160715)) of GeMAPS feature set, both related to phonemic and prosodic aspects. For dogs, we cannot state the internal causes in this paper as we don’t master the vocal expressions patterns of them till now. But this interesting topic has its value to be discussed in our future work.

**Reviewer#2:**

**Q: English titles don’t imply English-speaking.**

A: During our searching for videos, we have tried to get them from corresponding language environments by setting ip address in relative nations, checking titles and checking captions at the same time. We agree that containing English titles and captions doesn’t ensure that the video is certain to come from English-speaking families. Therefore, we have sampled 1,000 English videos in our dataset to check, a 93.2% accuracy is revealed.

**Reviewer#3:**

**Q: Concerns for interpreting the results.**

A: In deriving our results, we have tried our best to eliminate other factors influencing the barks of dogs in our approach and got the results shown in the paper. In our paper, our statements are not conclusive, but certain influence of different ambient languages are revealed.

**Q: The bold claim to say “dogs’ spoken language”.**

A: We sincerely thank you for pointing out this issue. It’s too bold for us to use “spoken language” in some parts of our paper. “Vocal expressions” is a more proper term. We will revise that in the revised version of paper.