# Rebuttal

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.

**General 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 possible 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 will make it clear in our next version.

**Q: The influence of other confronting factors like body size, age, sex, mood, and 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 emphasized that these factors are not related to the cultures of different nations. As we sampled our dataset from a wide range of sources for dogs from both environments, the distributions of these factors should not vary between the two language environments. At the same time, noises both exist in two environments. We have tried our best to reduce impact by enlarging the population of dogs in our dataset. However, the influence coming from these factors is interesting for sure, and we may take research on that in future work.

**Reviewer1:**

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

A: The variable of "unvoiced segment length" is defined in the original GeMAPS paper[1], both related to phonemic and prosodic aspects. For dogs, we cannot state the internal causes in this paper as we don't master their vocal expression patterns till now. But this interesting topic has its value to be discussed in our future work.

**Reviewer2:**

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

A: During our search for videos, we tried to get them from corresponding language environments by setting the IP address in related nations, checking titles, and checking captions. 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, and a 93.2% accuracy is revealed.

**Reviewer3**:

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

A: We applied official definitions provided by GeMAPS for acoustic features and used SHAP to explain their importance. The whole procedure is controlled and followed previous conventions. In our paper, our statements are not conclusive, but the certain influence of different ambient languages is 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 our next version of the paper.

[1]Eyben F, Scherer K R, Schuller B W, et al. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing[J]. IEEE transactions on affective computing, 2015, 7(2): 190-202.