

Innovation in robotic hearing

Volume 9 Issue 2 - 2023

Dan Wang

Assistant Professor, Department of Materials Science and Engineering, China

Correspondence: Dan Wang, Assistant Professor, Department of Materials Science and Engineering, Beijing Institute of Technology, Haidian District, Beijing, China 100811, Tel +86 13716799367, Email dan.wan@bit.edu.cn

Received: May 29, 2023 | **Published:** May 30, 2023

Editorial

In a world where smarty-pants systems are elbowing their way into our daily routines, the ability for these systems to accurately distinguish between environmental noise and human speech is of vital importance. The work of this research team, published recently,¹ has taken a stride in this regard.

This study zeroes in on a usual, but tricky situation: a friendly robot interacting with adults while a TV buzzes in the background. Despite this noise being everywhere in many households, it presents a unique challenge for robots designed to react to human speech. The investigators have found a solution using a convolutional neural network (CNN), a model typically employed in image processing, to distinguish between human voices and TV sound.

The team's approach to data collection is noteworthy. They have gathered and analyzed sounds from 18 different households, ensuring the findings of this study are rooted in real-world scenarios rather than idealized conditions. This authentic approach to data collection ensures their research reflects the messy realities of everyday life, enhancing its practical relevance. The authors used spectrograms and mel-spectrograms to convert audio data into a format that a CNN can process. Their CNN model achieved results with reasonable accuracy. Considering the variety and complexity of the sounds encountered in a typical home environment, this is a significant achievement. They also recognized that each household contributes different types of data and audio patterns, which would influence the CNN's performance. Their subsequent analysis revealed an unexpected result- a one-way influence between data from different households. This study lays a foundation for the development of robots that can better navigate and operate in the complex and noisy environments of our daily lives.

Furthermore, the research introduces new questions for exploration. What characteristics does the CNN learn to distinguish TV sounds from human speech? How can these insights be used to optimize the neural network for better detection and real-time recognition? As we push further into a future filled with smart devices and artificial intelligence, these are the kinds of questions that will help us to build systems that truly understand and interact with users.

In conclusion, this research study is a testament to the strides we are making in the field of intelligent systems, and its results are meaningful for what they have achieved and for new avenues of research they have opened up. As we continue to improve our AI systems, more meticulous studies are what we need.

Acknowledgements

None.

Conflicts of interest

Author declares there is no conflict of interest.

References

1. Figueroa D, Nishio S, Yamazaki R, et al. Improving voice detection in real life scenarios: differentiating television and human speech at older adults' houses. *Int Rob Auto J.* 2023;9(1):8-13.