An $YMnO_3$ Single Crystal-Based In-Materio Physical Reservoir Computing Device for Voice Recognition

Kyushu Inst. Tech. Neumorph Center[1], Kyushu Inst. Tech. Mater. Sci. Eng.[2],
Kyushu Inst. Tech. Hum. Intel. Sys.[3]
°Muzhen Xu[1], Kyoka Furuta[2], Ahmet Karacali[3], Yuki Umezaki[2],
Yuki Usami[1,3], Yoichi Horibe[1,2], Hirofumi Tanaka[1,3]
E-mail: xumuzhen@brain.kyutech.ac.jp

[Introduction] Physical reservoir computing (PRC) is an innovative computational paradigm that leverages intrinsic nonlinearity of physical systems to efficiently perform complex tasks. Yttrium manganese oxide ($YMnO_3$), a unique ferroelectric material, features a network of semiconductive domains and walls analogous to a reservoir layer.[1] This study aims to evaluate the voice recognition performance of $YMnO_3$ as an in-material PRC device.

[Methods] The surface of $YMnO_3$ crystallites was polished perpendicularly to the crystallographic c-axis (namely $YMO_{\perp}$), and a 16-electrode array was deposited on both the top and bottom sides. A free-spoke-digit-data dataset containing voice recordings of the numbers zero to nine, pronounced 47 times each by six speakers (George, Lucas, Jackson, Theo, Nicolas, and Yweweler), was used for voice recognition. This dataset was augmented 15 times using a sulfonated polyaniline network[2], followed by rectification, downsampling, and normalization (range: 0-5) using Python software. The pretreated voice signals were then amplified by three times (Thurlby Thandar Instruments, WA301 Wide Band Amplifier, 30V pk-pk) and applied to the $YMnO_3$ PRC system as time-series bias voltages using LabVIEW software. 31 output signals from the device were recorded simultaneously with a sampling rate of 1000 points/s. For voice recognition, all output signals were labelled as real numbers, a one-hot vector was used as the target, and ridge regression was applied for classification using Python software.

[Results and Discussion] Unlike conventional methods such as Mel frequency cepstral coefficients (MFCC) or cochleagram, this study employed a simplified preprocessing approach for raw voice digits. As shown in Fig. 1(a), the raw data were augmented 15-fold to increase the database size. The augmented data were then rectified to the positive side and downsampled from 2000 to 100 points. Finally, the envelope of the augmented data was derived to create a simplified representation of the voice digit, which was normalised and used as the input. This streamlined preprocessing approach could potentially be implemented in hardware, significantly enhancing energy efficiency compared to software-based methods. It demonstrated impressive recognition accuracies across different digits and speakers (Fig. 1(b-c)), achieving up to 75% accuracy for digits and 98% accuracy for speakers. These results highlight the potential of $YMnO_3$-based PRC for practical applications in speech recognition, offering both high performance and energy efficiency.

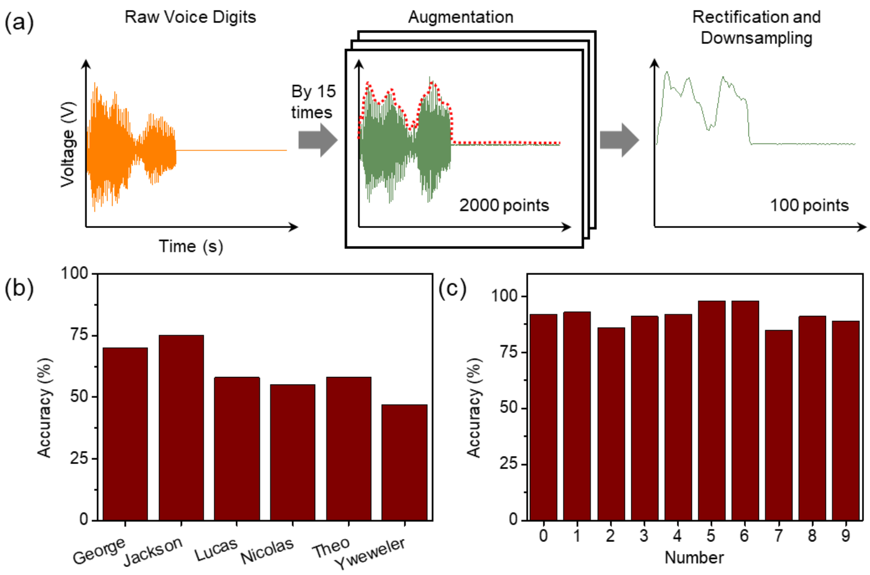

Fig. 1 (a) Schematic representation of the voice pretreatment procedure. (b-c) Recognition accuracy of $YMO_{\perp}$ device for (b) ten digits spoken by six speakers and (c) six speakers articulating ten digits.

[1] Choi, T., et al. *Nature Mater.* 9, 253–258 (2010).
[2] Usami, Y., et al. *Adv. Mater.* 33, e2102688, (2021).