A FREQUENCY RESPONSE NORMALIZATION METHOD BASED ON DEEP NEURAL NETWORKSFOR IOT DEVICES

[1]DEOKGYU YUN, [2]JAEGYU CHOI, [3]HANNA LEE, [4]SEUNG HO CHOI

[1]Dept. of Electronic Engineering, Seoul National University of Science and Technology
[2,3,4]Dept. of Electronic and IT Media Engineering, Seoul National University of Science and Technology
E-mail: [1]deokkyuyun@gmail.com, [2]gyuuu_@naver.com,[3]hannha9656@gmail.com, [4]shchoi@snut.ac.kr

Abstract - This paper presents a novel frequency response normalization method for IoT(Internet of Things) devices, which is based on deep neural networks. When connecting the sounds acquired by different smart phones, unnatural soundsare generated, which is mainly due to different frequency responses for each smart phone. To solve this problem, we need to normalize the frequency responses of smart phones. In this study, we propose a normalization method using deep neural network. The input of the neural network is the spectrum of the smart phone sound to be processed and the output is the ratio of input to the spectrum of the reference smart phone sound. Experimental results show that when the acoustic signals acquired by different smart phones are connected, the naturalness of the sound is improved through objective and subjective evaluation.

Keywords - Frequency normalization, IoT(Internet of Things), Virtual reality, User created contents, Deep neural network.

I. INTRODUCTION

Frequency responses of IoT(Internet of Things) devices can be different. Therefore, frequency response normalization is necessary for the system using IoT devices. Especially, in order to provide a realistic media suitable for virtual reality using user created contents, the video and audio signals captured by their respective smart phones have different characteristics. In this case, it is necessary not only to match the various videos received at different positions [1] but also to synchronize and normalize the different audio signals. In this paper, to solve this problem, we propose a method of frequency normalization of the acoustic signals acquired by different smart phones using a deep neural network (DNN). The DNN is trained to output the ratio of the spectrum of an acoustic signal having a different frequency response to the spectrum of a specific smart phone model. Through this process, the sounds of different smart phones can be smoothly connected.

II. COMPARISON OF FREQUENCY RESPONSES

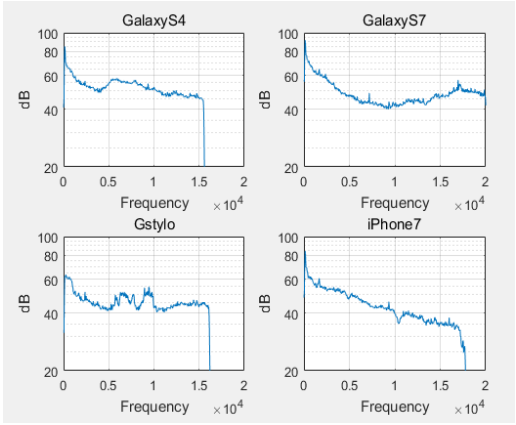

Fig.1Frequency response of each smartphone

First, to compare the frequency responses of different smartphone models, we analyzed the frequency response with a test signal whose frequency increases with time. Fig. 1 shows the frequency response of four models. As can be seen, the cut-off frequency and gain are different.In this study, considering the cutoff frequencies of other smartphones, we use iPhone 7 as a reference smartphone.

II. DNN-BASED FREQUENCYRESPONSENORMALIZATION

In the proposed method, the input of DNN is the spectrum of each smartphone model and the DNN is trained to output the ratio of the spectrum of the reference smartphone to that of the other smartphone. As the activation function, we use the ReLU (Rectified Linear Units) [2] at the input of the node and the drop out [3] that follows the activation function. In addition, ADAM (ADAmptive Moment Estimation) [4] is used for the learning algorithm. The overall structure is shown in Fig. 2. InFig.2, $X_{in}(i)$isthe spectral bin of a smartphone modeland $X_{ref}(i)$ is that of the reference one. The ratios are computed as in Eq. (1).

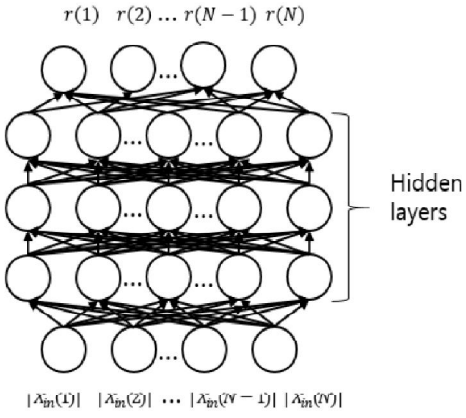

Fig.2 Structure of DNN

$$r(i) = \frac{|X_{ref(i)}|}{|X_{in(i)}|}, \; i = 1,2 \cdots N, \quad (1)$$

where $i$and $N$are frequency bin index andthe number of the frequency bins, respectively.The normalized spectrum of each smartphone is achieved by obtaining thespectral ratio estimates through the trained DNN and then multiplying the input spectrum by these ratios.

III. EXPERIMENT AND RESULT

First, we acquired various sounds for about 30 minutes with several smartphones of different models with the sampling frequency of 44.1 kHz. After synchronizing the signals, theshort-time spectra were obtained.In this experiment, the frame size was 20 ms with 50% overlap and the Hanning window was applied. The number of input and output nodes were 513 that corresponds to the number of frequency bins. The number of hidden layers of the neural network is three and the each hidden layer has 1024 nodes.

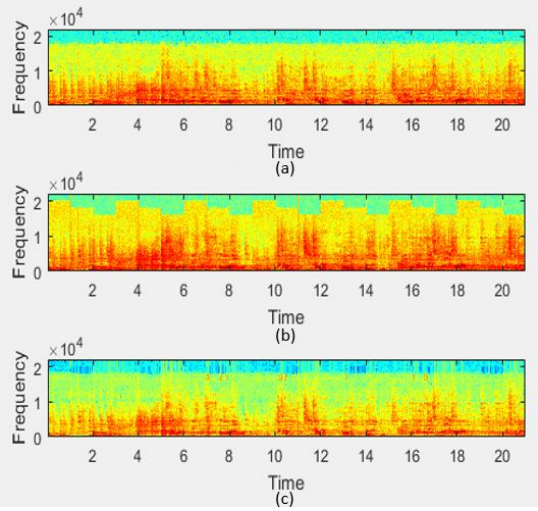

Fig.3Spectrogram examples for (a) reference(b) before normalization and (c) after normalization.

Fig. 3 shows the spectrogram examples of the signals acquired by the reference model (a)and of the different models at one second intervals (b), and of the normalized signals (c). As shown in the Figure, the spectrogram of the processed signal is more similar to the reference one. From informal listening tests, the processed sounds weremore natural than those before normalization. In addition, log spectral distance (LSD) was calculated to compare the results. The LSD between the signals of the reference andthe other models wasreduced from 8.19 dB to 6.33dB after processing.

CONCLUSION

In this paper, we developeda DNN-based frequency normalization methodthat compensates the difference of frequency responsesof IoT devices. From the spectrograms and LSD distances in the experiments,we showed that the normalization method is effective. Therefore, it is possible to obtain a natural sound whendifferent sounds are connected.

ACKNOWLEDGMENTS

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2017-2016-0-00311) supervised by the IITP(Institute for Information & communications Technology Promotion).

RFERENCES

[1] G.E. Christensen and H.J. Johnson, “Consistent image registration,”IEEE Transactions on Medical Imaging, pp.568 - 582, 2001.

[2] V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” in Proc. 27th Int. Conf. Machine Learning, pp. 807-814, 2010.

[3] Vu Pham, ThéodoreBluche, Christopher Kermorvant, and JérômeLouradour, “Dropout improves recurrent neural networks for handwriting recognition,” Proc.Frontiers in Handwriting Recognition (ICFHR), pp. 285–290, 2014.

[4] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

★★★