

# Text-Independent Speaker Identification Using Deep Learning Model of Convolution Neural Network

Supaporn Bunrit, Thuttaphol Inkian, Nittaya Kerdprasop, and Kittisak Kerdprasop

**Abstract**—Speaker recognition approach can be categorized into speaker identification and speaker verification. These two subfields have a bit varied in definition from domain usage. If we have a voice input, the goal of speaker verification is for authentication by determining an answer from a question: “is the voice someone’s voice?” For speaker identification, will try to find an answer: “the voice is whose voice?” It can be thought that verification is a special case of open-set identification. In this work, deep learning model using a convolution neural network (CNN) for speaker identification is proposed. The voice input to the method is no constrained on the words the speaker speaks. That means it is in a form of text-independent of which more difficult than text-dependent system. By the method, each 2 seconds of the speaker voice is transform to a spectrogram image and input to the generated CNN model training from scratch. The proposed CNN based method is compared to the classic Mel-frequency cepstral coefficients (MFCCs) based featured extraction method classified by support vector machine (SVM). Where, up to date, MFCC is the most popular feature extracted method for audio and speech signal. Our proposed method that the spectrogram image is used as an input is also compared to a case when image of raw signal wave is employed to the CNN model. Experiments are conducted on the speech from five speakers speak in Thai language of which voices are extracted from YouTube. It reveals the proposed CNN based method trains on spectrogram image of voice is the best compared to the other two methods. The average classification results of the testing set by the proposed method is 95.83%. For MFCC based method is 91.26% and for CNN model trained on image of raw signal wave is only 49.77%. The proposed method is very efficient when only short utterance of voice is used as an input.

**Index Terms**—Convolution neural network (CNN), deep learning, speaker recognition, speaker identification, text-independent.

## I. INTRODUCTION

Technology of voice recognition can be roughly divided into two sub-areas, which are speech recognition and speaker recognition [1]. Speech recognition is an approach for analyzing the contents of the words/speech that a speaker speaks. Each of the speech recognition system use many of algorithms to convert the sound waves into useful data for processing which is then interpreted by the machine. These

systems then produce an output generated in the form of text to be used [2]. Whereas, speaker recognition is the method used to identify whom the speaker speak such words/speech. The techniques for speaker recognition attempt to cover the different aspects for recognizing persons from their voice. Because each speaker has his or her characteristic manner of speaking, including the use of a particular accent, rhythm, intonation style, pronunciation pattern, choice of vocabulary and so on [3]. Speaker recognition systems can be divided into text-dependent and text-independent. In text-dependent systems suited for cooperative users, where the recognition phrases are fixed, or known beforehand [4]. In text-independent systems, there are no constraints on the words that the speakers are allowed to speak. Thus, the reference (what are spoken in training) and the test (what are uttered in actual use) utterances may have completely different content, and the recognition system must consider this phonetic mismatch. Text-independent recognition is the much more challenging of the two tasks [3]. Moreover, in real life, text-independent systems are more commercially attractive than text-dependent systems because it is harder to mimic an unknown phrase than a known one [5]

In this work, we focus on text-independent of speaker recognition, which can be roughly divided into two parts: speaker identification and speaker verification [6]. These two subfields have a bit varied in definition from domain usage. The goal of speaker verification is for authentication by determining an answer from a question: “is the voice someone’s voice?” For speaker identification, will try to find an answer: “the voice is whose voice?” It can be thought that verification is a special case of open-set identification. Where, speaker identification is the process of identifying the speaker from a given utterance by comparing the voice biometrics of the utterance with those utterance models stored beforehand [7]. In our study, the method for text-independent speaker identification is proposed.

The classical Mel-frequency cepstral coefficients (MFCCs) method is likely the most popular feature extraction strategy used to date [6] for audio and speech signal. From literature, most of the state-of-the-art speaker identification systems use MFCC as feature extractor, and then feed these features to GMM-based approaches for creating speaker models for identification [7]. Nevertheless, the researchers are primary proposed the techniques suitable for a large amount of speech in training and test [8]. Considering for the short utterances of speech in speaker identification, as a result, is a challenging domain. For our proposed work, CNN based method using spectrogram image from short utterances of speech are investigated. The method experiment compare to the classic feature extracted of MFCC and CNN that use signal wave image as an input.

Manuscript received August 25, 2018; revised March 19, 2019. This work was supported by grants from Suranaree University of Technology (SUT), Thailand.

The authors are with the School of Computer Engineering, SUT, 111 University Avenue, Muang, Nakhon Ratchasima 30000, Thailand. (corresponding author: S. Bunrit; tel.: +66944961244; e-mail: sbunrit@sut.ac.th, thuttapholti@gmail.com, nittaya@sut.ac.th, kerdpras@sut.ac.th).

## II. LITERATURE REVIEWS

By an advancement of artificial intelligence technology, there emerge various applications needed the system of speaker identification. The researchers in a field then dramatically investigate novel technologies for speaker identification as a new research direction. The most emphasis process for speaker identification is the featured extraction process. The features extracted from each proposed method always affect the performance gained by the method. In literatures, pure Mel-Frequency Cepstral Coefficients (MFCCs) based feature extraction approaches have been used more than any other approaches. Furthermore, other MFCC variations, such as MFCC fusion and cleansing approaches, are proven to be very popular as well [7].

Normally, the methods for speaker identification task can be divided into two mainly processes, which are the training process and the testing (identification or matching) process. The training process can also be separated into 4 stages as follow:

- 1) Inputting the speech signal
- 2) Pre-processing of the speech
- 3) Normalization
- 4) Feature extraction

As mentions previously, feature extraction is the most significant stage for speaker identification. As a result, most researches tried to explore the varieties of techniques for extracting the feature from speech data aims at identifying of each speaker. Because MFCCs based feature extraction approaches have been tentatively used more than any others did, therefore, we can category the methods for feature extraction of speaker identification task in to MFCC-based approach and non-MFCC-based approach.

For MFCC-based approach, the filter banks are designed in a manner to operate in a similar way to the human auditory frequency perception [9]. Many fusions of MFCC-based featured have been continuously studied [10]-[12]. By combining two different set of features from MFCCs and Perceptual Linear Predictive Coefficients (PLPC) using ensemble classifiers in conjunction with principal component transformation can significantly improve the performance of the MFCC-GMM speaker recognition systems in [10]. Wang and Johnson [11] also investigated MFCC features combined with Residual Phase Cepstrum Coefficients (RPCC). It offered significant overall improvement to the robustness and accuracy of speaker identification tasks. Recently, Ma, Yu, Tan, and Guo [12] used MFCC features incorporated to histogram transform feature for text-independent speaker identification task.

A group of non MFCC-based approach employed other methods to extract the features of speech for speaker identification. Example of such non-MFCC features are Auditory-based time frequency Transform (AT) proposed in [13]. Mean Hilbert envelope coefficient (MHEC) is a feature that studied in a work of Godin *et al.*, [14] and a work of Sadjadi and Hansen [15]. Power-normalized cepstral coefficient (PNCC) feature was proposed in [15] and [16]. Where Feature extraction method using Fisher Vector (FV) also investigated by Jiang *et al.*, [17]. Nevertheless, Lukic *et al.*, [18], explored a novel deep learning technique. They proposed deep learning based on CNNs model and that verified the model by using connected speech samples

provided by TIMIT.

### A. Convolution Neural Network (CNN)

CNN is one particular neural network model in deep learning technique. It is well known in machine learning and computer vision communities since ImageNet Large-Scale Visual Recognition Challenge (ILSVRC 2012) when CNN architecture of AlexNet [19] by Krizhevsky *et al.*, won the competition. The name tells this network use the convolution operation as the key processing operation. CNN network came from three of concepts, which are sparse interaction, parameter sharing, and equivariant representation [20]. Such concepts transform to the network configuration shows in Fig. 1. In order to learn and extract the features from input, three of principle stages are used which are convolution stage follow by nonlinearity stage using ReLU (Rectified Linear Unit) function and pooling stage. Many of these principle stages are consecutively used as layer-by-layer in the network for automatically extracting the deep features. The beginning layers to the intermediated layers are used aiming at extracting the features by the learning process. Such learned features are then sending to the layers near the end of the network for the classification process. Where, the Fully Connected (FC) networks as the traditional neural network manner are always used. Finally, softmax function is applied for the last layer in order to gain the output in probability fashion.

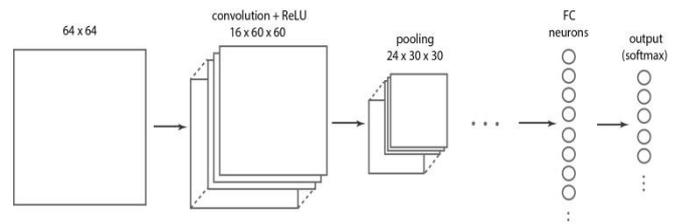


Fig. 1. Principle stages of convolution neural network model.

For each convolution stage/layer in CNN, many filters of size  $k \times k \times D$  are employed to convolute with the incoming input as shown in Fig. 2. Filter is applied once at a time to convolute with the input in a form of sliding windows. The convolution result from one filter is one of the feature map output. As a result, when  $N$  filters are applied, the whole output from the convolution stage is the stack of  $N$  feature maps as depict on the right side of Fig. 2.

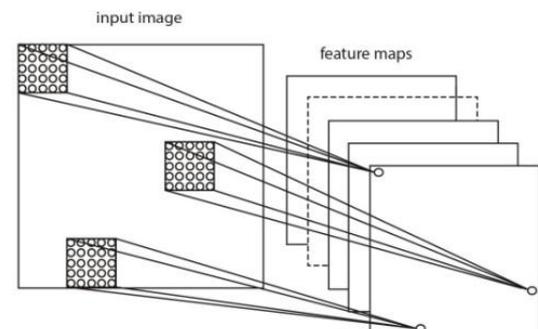
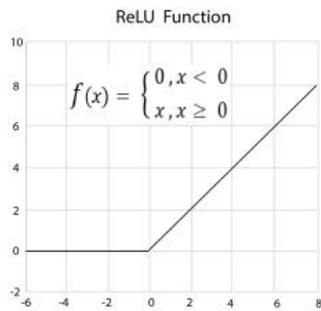


Fig. 2. Convolution operation in CNN.

Feature maps result from the convolution stage will pass through a non-linear Rectified Linear Unit (ReLU) function in order to extract the non-linear property of the features. ReLU function is shown in Fig. 3(a). The function converts all the negative values to zero where keeps the others as the original.

Example of values after applied ReLU function also shows in Fig. 3(b).

The stage of pooling in a network uses for subsampling the data to the previous stage feature maps output. After the pooling stage, as a result, the dimension of width (W) and height (H) of the feature maps will decrease. Differences of the subsampling size can be applied for the pooling stage. For example, when the pooling size of 3×3 windows is used, the only one value from nine values in the windows is selected as the subsampling value. Therefore, max pooling, average pooling, or any others pooling types can be applied to selected such one subsampling value. Fig. 4 shows an example when the average pooling and max pooling are employed to the same locations of 3×3 windows size.



(a). Retified Linear Unit (ReLU) function

17	80	-115	8	23	→	17	80	0	8	23
-99	55	70	35	150		0	55	70	35	150
25	-9	205	5	88		25	0	205	5	88
10	75	5	-205	3		10	75	5	0	3
105	-12	60	2	-77		105	0	60	2	0

(b). Example of values in feature map after applied ReLU function.  
Fig. 3. Nonlinear operation by ReLU function.

20	8	11	5	17	23	Average Pooling	18	20
8	10	19	21	8	11		14	13
12	3	17	9	18	7	Max Pooling	20	23
5	10	3	7	1	15		17	19
11	9	11	2	19	3			
17	12	4	14	8	7			

Fig. 4. Pooling operations always use in CNN model.

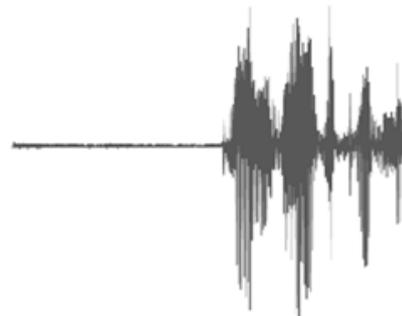
In CNN model, many layers consisting of the previous mention principle stages, which are convolution, nonlinearity by ReLU, and pooling are consecutively arranged to form a network in feature learning process. Some other stages may include such as the stage of normalization and dropout. Then, the extracted features that learned from the feature learning process will be arranged in a vector form and some of FC layers the same as the traditional multi-layer perceptron (MLP) are used in the layers before softmax layer. Finally, for the layer before output, softmax function is applied in order to transform the output values of the network in term of probability. Softmax function is defined by (1).

$$\text{SoftMax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad , j = 1, 2, \dots, K \quad (1)$$

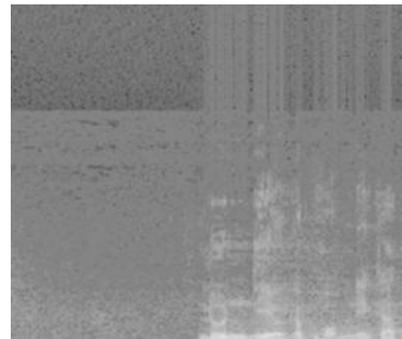
where  $z_j$  is an output of each  $j$ ,  $e^{z_j}$  is the exponential value of  $z_j$ , and  $k$  is the component of vector  $Z$ .

### B. Spectrogram Image

A spectrogram is a visual representation of the spectrum of frequencies of sound or other signal as they vary with time. Spectrograms are sometimes called sonographs, voiceprints, or voicegrams. When the data is represented in a 3D plot they may be called waterfalls [21]. Figure 5(b) depicts a 2D example of spectrogram in gray scale image. It is a spectrogram transformed from the signal wave in 2 seconds of speech depicted in Fig. 5(a). It can be inferred that spectrogram is an image of its spectrum.



(a). Image of raw signal wave of speech in 2 seconds.



(b). Spectrogram image of raw signal wave in a.  
Fig. 5. Spectrogram of speech signal.

A spectrogram image can be created by applying a Fast Fourier Transform (FFT) to the recorded signal of speech. The FFT can analyze the signal by essentially separating the frequencies and amplitudes of its component simplex waves. The result can then be displayed visually, with degrees of amplitude at various frequencies by time. Spectrogram image will represent the level the energy from light-to-dark. In case it represent color is white or nearly white, its means there is no or small energy. Oppositely, If they has a lot of energy, it represent color is black or nearly black.

### C. Mel-Frequency Cepstral Coefficients (MFCCs)

The processes of MFCCs based feature extracted method are shown in Fig. 6. The method starts by the normalization and preprocessing of the speech or acoustic signal if it is needed. Then, the time analysis is performed by employing framing and windowing operation before the FFT process. After that, Mel-scale filter bank is applying and wrapping into the logarithm scale. Discrete Cosine Transform (DCT) is then used as a process before calculating the Cepstral Mean Subtraction (CMS). Such CMS coefficients are the features extracted by MFCCs method. That means, MFCCs based feature extracted method used frequency analysis based on

progressing the speech frame through filter bank [7]. The result from the method is the extracted features in a form of MFCCs coefficients. These MFCCs coefficients, as a result, can be further employed for analyzing or classifying by any purpose.

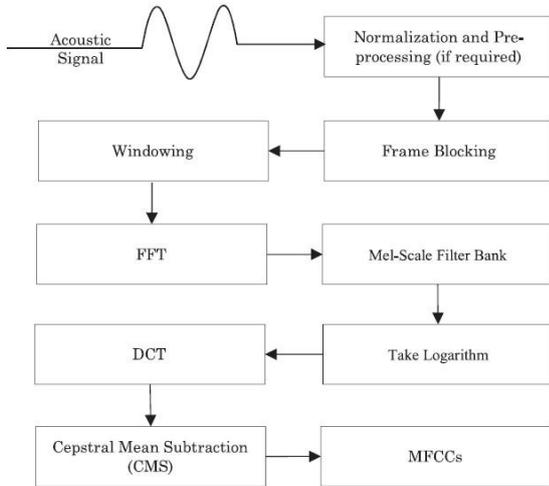


Fig. 6. MFCCs feature extraction method [7].

### III. PROPOSED WORK

We use CNN based training from scratch in our proposed work. Therefore, the appropriated CNN model for the studied data set is generated using an empirical experiment. We implement the model by Matlab. The structure of a CNN model proposed for Thai speaker identification shows in Table I. The model train on 1000 epoch with learning rate of 0.001. The stochastic gradient descent with momentum (SGDM) optimization algorithm is employed. The momentum term is set to 0.9 and the regularization term is 0.5.

TABLE I: STRUCTURE OF THE PROPOSED CNN MODEL

Layer No.	Layer Name	Detail
1	Input	187×212×1 (gray scale)
2	convolution	3×3, 8 filters, pad 1
3	batchNormalization	
4	ReLU	
5	maxPooling	2×2, stride 2
6	convolution	3×3, 16 filters, pad 1
7	batchNormalization	
8	ReLU	
9	maxPooling	2×2, stride 2
10	convolution	3×3, 32 filters, pad 1
11	batchNormalization	
12	ReLU	
13	fullyConnected	5 neurons
14	softMax	
15	classificationOutput	5 classes

As demonstrate in Table I, Our proposed CNN model train from scratch employs only three of the convolution layers, there are in layer number 2, 6, and 10, respectively. Pooling layers with max pooling are used in layer number 5 and 9. Where, batch normalization layer is inserted after all convolution layers. ReLU layer is then also followed all

batch normalization layers. We use only one of fully connected layer, which is a layer before softmax and output layer.

The detail of each layer is also explained on the right most column of Table I. Input to a network is a gray scale image of spectrogram or a gray scale image of raw signal wave. Both are in sized of 187×212. The first convolution layer uses 8 filters of size 3×3. The second convolution layer uses 16 filters of size 3×3, and the third convolution layer uses 32 filters of size 3×3. All pooling layers use the subsampling in a windows size of 2×2, and striding by two positions. In this work, we experiment on a dataset from 5 speakers speak in Thai language, there, an output from the classification output layer is set to 5 for 5 classes we need to classify.

## IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

### A. Data Set Preparation

Our work focuses the study to the speakers that used Thai language in social media input. We, therefore, used five of Thai speaker's voice in .wav format from YouTube. Then, Audacity software is used to convert such audio files into nomo files. Finally, we consecutively cut each file into 100 samples of 2 seconds each by Matlab. For each speaker, 80 samples are for training and the remaining 20 sample are for testing.

For MFCC method, MFCC coefficients are extracted as the features from each sample and SVM is used as the classifier. For a method that use raw signal wave image to CNN, we plot signal wave of each sample and save it as an image file in JPEG format for training and testing with CNN model. An example of the signal wave image is already shown in Fig. 5(a). In addition, for our proposed method, each signal wave sample is transform to spectrogram image. Such images are then use for training and testing with the appropriated generated CNN model. An example of the spectrogram image is also shown previously in Fig. 5(b).

### B. Experimental Results

Table II shows the accuracy result from the classification when the proposed method is compared to MFCC method and CNN of signal wave method. The experiments conduct on 50 times to the testing set for each method. Maximum and average of the classification accuracy show for comparison. It reveals the proposed CNN based method using spectrogram image get the highest average accuracy of 95.83%. When signal wave image is used as an input to the CNN model, it exposes very bad result, only 49.77% on an average accuracy. In case of the traditional MFCC based method, the average accuracy is less than the proposed method around 5%.

TABLE II: COMPARE THE CLASSIFICATION ACCURACY FROM EACH METHOD

Method	Accuracy (%)	
	Max	Average
MFCC	91.30	91.26
CNN of Signal Wave	54.00	49.77
CNN of Spectrogram (Proposed)	<b>99.00</b>	<b>95.83</b>

Fig. 7 shows the classification result from the proposed

CNN based method in term of confusion matrix. It is a performance result from the testing set. Each speaker (named as Sp1 to Sp5 in a figure) has 20 samples for testing. Entirely, the average classification result is 95.83% (in a matrix shows 96.0% cause by ceiling).

Output Speaker	Sp1	20 20.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	Sp2	0 0.0%	19 19.0%	2 2.0%	0 0.0%	0 0.0%	90.5% 9.5%
	Sp3	0 0.0%	0 0.0%	18 18.0%	0 0.0%	0 0.0%	100% 0.0%
	Sp4	0 0.0%	1 1.0%	0 0.0%	20 20.0%	1 1.0%	90.9% 9.1%
	Sp5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	19 19.0%	100% 0.0%
		100% 0.0%	95.0% 5.0%	90.0% 10.0%	100% 0.0%	95.0% 5.0%	<b>96.0%</b> <b>4.0%</b>
	Sp1	Sp2	Sp3	Sp4	Sp5		
	<b>Target Speaker</b>						

Fig. 7. The classification results from the proposed method.

### C. Discussions

By the proposed CNN based method that creating the CNN models from scratch, a model uses a spectrogram image as an input reveals a very attractive result to the studied dataset. That means, transforming the raw signal data into the spectrogram images and learn them by the suitable creating CNN model could be extracted the variance distinct features of the speakers. The proposed model, therefore, exposes the high accuracy result compared to the traditional method. Most researches of speaker identification were studied on the dataset that speakers speak on English language. Moreover, in literatures, none of work related CNN based method studied for Thai speaker identification. In this work, we also focus on the two difficult challenges in speaker recognition, which are a challenge when text-independent of speech is used and a challenge when only the short utterances of speech are considered. For the future research directions, we are still interesting in deep learning models and Thai speaker identification. That may investigate on the transfer-learning scenario when some of the CNN pre-trained models are applied to the studied dataset. In a case when consider a very short utterances of speech, another model of deep learning named deep autoencoder is also interesting.

### V. CONCLUSION

CNN based method by training from scratch for text-independent speaker identification is proposed in this work. Each sample of signal wave is transformed to spectrogram and use as a gray scale input image to the network. Our proposed method that the spectrogram image use as an input also compare to a case when image of raw signal wave is employed to the CNN model and MFCC based method. Experiments are conducted on five speakers speak in Thai language of which voices are extracted from YouTube.

It reveals the proposed CNN based method trains on spectrogram image of voice is the best compared the other two methods. The average classification results of the testing set by the proposed method is 95.83% of accuracy. For MFCC based method is 91.26% and for CNN trained on image of raw signal wave is only 49.77%. The proposed method is very efficient for text-independent approach where only short utterance of voice is needed as an input.

### REFERENCES

- [1] L. Rabiner L, B. H. Juang, *Fundamentals of Speech Recognition*, Englewood Cliffs, NJ: Prentice - Hall International, 1993.
- [2] S. V. Ault, R. J. Perez, C. A. Kimble, and J. Wang, "On speech recognition algorithms," *International Journal of Machine Learning and Computing*, vol. 8, no. 6, pp. 518-523, 2018.
- [3] T. Kinnunen, and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech Communication*, vol. 52, pp. 12-40, 2010.
- [4] M. He bert, "Text-dependent speaker recognition," in *Springer Handbook of Speech Processing*, J. Benesty, M. Sondhi, Y. Huang, Eds. Springer-Verlag, Heidelberg, 2008, pp. 743-762.
- [5] N. Singh, R. A. Khan, and R. Shree, "Application of speaker recognition," *Procedia Engineering*, vol. 38, pp. 3122-3126, 2012.
- [6] J-C. Liu, F-Y. Leu, G-L. Lin, and H. Susanto, "An MFCC-based text-independent speaker identification system for access control," *Concurrency Computat: Pract Exper.*, vol. 30, no. 2, Article ID e4255, 2018.
- [7] S. S. Tirumala, S. R. Shahamiri, A. S. Garhwal, and R. Wang, "Speaker identification features extraction methods: A systematic review," *Expert Systems with Applications*, vol. 90, pp. 250-271, 2017.
- [8] A. Poddar, M. Sahidullah, and G. Saha, "Speaker verification with short utterances: a review of challenges, trends and opportunities" *IET Biometrics*, vol. 7, no. 2, pp. 91-101, 2017.
- [9] Z. Ma and A. Leijon, "Super-Dirichlet mixture models using differential line spectral frequencies for text-independent speaker identification," *INTERSPEECH*, pp. 2360-2363, 2011.
- [10] S. Bose, A. Pal, A. Mukherjee, and D. Das, "Robust speaker identification using fusion of features and classifiers," *International Journal of Machine Learning and Computing*, vol. 7, no. 5, pp. 133-138, 2017.
- [11] J. Wang, and M. T. Johnson, "Physiologically-motivated feature extraction for speaker identification," in *Proc. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, 2014, pp. 1690-1694.
- [12] Z. Ma, H. Yu, Z. H. Tan, and J. Guo, "Text-independent speaker identification using the histogram transform model," *IEEE Access*, vol. 4, pp. 9733-9739, 2016.
- [13] O. Plchot, S. Matsoukas, P. Matejka, N. Dehak, J. Z. Ma, and S. Cumani, "Developing a speaker identification system for the DARPA RATS project," in *Proc. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, 2013, pp. 6768-6772.
- [14] K. W. Godin, S. O. Sadjadi, and J. H. L. Hansen, "Impact of noise reduction and spectrum estimation on noise robust speaker identification," *INTERSPEECH*, 2013, pp. 3656-3660.
- [15] S. O. Sadjadi, and J. H. L. Hansen, "Mean Hilbert envelope coefficients (MHEC) for robust speaker and language identification," *Speech Communication*, vol. 72, pp. 138-148, 2015
- [16] V. Mitra, M. McLaren, H. Franco, M. Graciarena, M. and N. Scheffer, "Modulation features for noise robust speaker identification," *INTERSPEECH*, pp. 3703-3707, 2013.
- [17] S. Jiang, H. Frigui, and A. W. Calhoun, "Speaker identification in medical simulation data using fisher vector representation," in *Proc. 2015 IEEE 14th International Conference on Machine Learning and Applications*, Miami, FL, USA, 2015, pp. 197-201.
- [18] Y. Lukic, C. Vogt, O. Dürr, and T. Stadelmann, "Speaker identification and clustering using convolutional neural networks," in *Proc. 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing*, 2016, pp. 1-6.
- [19] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *NIPS 2012*, pp. 1106-1114, 2012.
- [20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
- [21] Spectrogram. [Online]. Available: <https://en.wikipedia.org/wiki/Spectrogram>



**S. Bunrit** is a lecturer with computer engineering school, SUT. She received her bachelor degree in science (mathematics) from Kasetsart University, Thailand, in 1997, master degree in science (computer science) from Chulalongkorn University, Thailand, in 2001. Her research of interest includes artificial neural network, deep learning, machine learning, digital image processing, computer vision, and time series analysis.



**N. Kerdprasop** is an associate professor with Computer Engineering School, SUT. She received her bachelor degree in radiation techniques from Mahidol University, Thailand, in 1985, master degree in computer science from the Prince of Songkla University, Thailand, in 1991 and doctoral degree in computer science from Nova Southeastern University, U.S.A, in 1999. Her research of interest includes data mining, artificial intelligence, and intelligent databases.



**T. Inkian** is a bachelor's degree student with the school of computer engineering, SUT. His research of interest includes data sciences, artificial neural network and deep learning for vision system. currently, he also works as an integration tester of a software development company in Thailand.



**K. Kerdprasop** is an associate professor and chair of the School of Computer Engineering, SUT. He received his bachelor degree in mathematics from Srinakarinwirot University, Thailand, in 1986, master degree in computer science from the Prince of Songkla University, Thailand, in 1991 and doctoral degree in computer science from Nova Southeastern University, U.S.A., in 1999. His current research includes data mining, artificial intelligence, computational statistics.