Scalogram vs Spectrogram as Speech Representation Inputs for Speech Emotion Recognition Using CNN

Marc Dominic Enriquez
University of the Philippines
Digital Signal Processing Laboratory
Quezon City, Philippines
marc.dominic.enriquez@eee.upd.edu.ph

Crisron Rudolf Lucas
University College Dublin
Insight Research Centre
Dublin, Ireland
crisron.lucas@ucdconnect.ie

Angelina Aquino
Charles Darwin University
Northern Institute
Darwin, Australia
angelina.aquino@cdu.edu.au

2023 34th Irish Signals and Systems Conference (ISSC) | 979-8-3503-4057-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISSC59246.2023.10162085

***Abstract*—Speech Emotion Recognition (SER) focuses on understanding the human emotion in a given speech utterance using its acoustic and/or linguistic features. This paper presents a comparison between two speech representation inputs for SER: spectrograms and scalograms. Speech signals from four databases (Emo-DB, RAVDESS, SAVEE, and a mix of all three) were converted into each type of representation and were used to train variations of a convolutional neural network (CNN) VGG16 Model-3. Results show that the scalogram-based models have a higher mean f1-score compared to the spectrogram-based models; however, further analysis indicate that the difference is statistically insignificant at a 95% confidence level. In conclusion, spectrograms and scalograms have statistically the same performance on the systems presented.**

***Index Terms*—Spectrogram, Scalogram, SER, CNN, Fourier Transform, Wavelet Transform**

I. Introduction

Human emotion is an important layer in our communication and provides key information regarding a person's mental state. The same words and messages can easily change in meaning once the emotion is put into consideration [1]. Therefore, speech emotion recognition (SER) has been a subject of keen interest in the past couple of decades.

An SER system is commonly composed of three subsystems: data preprocessing, emotion feature extraction and selection, and emotion classification [2]. Studies have been exploring training neural networks to learn features from the input directly. These networks can operate on the speech signal itself but this often leads to a high number of inputs and high dimensionality of inputs, requiring more processing [3]. Hence, there is an incentive to perform neural network operations on a representation of the speech signal instead. Spectrograms have garnered a lot of attention and success, and a similar representation called a scalogram could have the same potential, but still lacks studies exploring its viability. Spectrograms and scalograms are similar in that they both are image representations of speech signals, but while the spectrogram is a product of the Fourier transform, the scalogram is produced when a wavelet transform is used. They both plot the frequency components of a speech signal, but the scalogram can do so with multiple resolutions.

This study was conducted to explore the viability and potential of scalograms as inputs to deep architectures. The study used the more established and widely used spectrogram-based inputs as a point for comparison. How the scalogram-based models perform in comparison to the Spectrogram-based models should provide valuable insight into the performance of scalogram-based inputs in this classification problem as a whole.

II. Related Work

This section of the paper will serve as a review of related works. It will first go over the different methods for SER and deep learning. It will be followed by a subsection explaining the popularity of CNNs and their advantages over other deep learning techniques. Then, the different potential inputs to CNNs will be discussed; this subsection will focus specifically on spectrograms, scalograms, and their features.

A. Deep Learning in SER

In the past decade, deep learning (DL) techniques have been finding use in different parts of SER [4]. These have shown improved results from older methods used such as Hidden Markov Models, Gaussian Mixture Models, and Decision Trees [5]. Furthermore, there are a few prominent choices when it comes to deep learning architectures such as Deep Neural Networks (DNN), Recurrent Neural Networks (RNN), Long Short-Term Memory Networks (LSTM), and Convolutional Neural Networks (CNN) [6]. A study by Sun et al. [7] utilized a DNN-based architecture to derive more discriminative features between easily confused emotions and achieved a higher recognition rate by 6.25% and 2.91% compared to traditional SVM and DNN-SVM-based methods respectively. A different study [3] opted to use Recurrent Neural Networks (RNN) instead. The study concluded that using a CNN with network layers based on LSTM, an artificial Recurrent Neural Network architecture, provides a viable SER algorithm that doesn't require any traditional hand-crafted

features. Finally, a study by Fayek et al. [6] demonstrated how CNN surpasses DNN and LSTM networks for SER. CNN and DNN were employed as static classifiers while the LSTM-RNN was regarded as a dynamic classifier. CNN yielded the best accuracy, followed by the DNN then lastly, the LSTM-RNN. Fayek proposes that this is because the static components in speech are more discriminative than its dynamic components. Furthermore, CNN is more robust to small variations in learning discriminative features.

B. Convolutional Neural Networks for SER

With these factors in mind, Convolutional Neural Networks have gained the most popularity for SER, even surpassing more complex models, for several reasons [4]. First, DL with CNNs can be used with raw sound recordings directly [8]. Second, CNNs can also be used with a processed representation of the speech signal [9]. The latter is a reason why CNNs are widely used for other classification problems such as computer vision [3]. Convolutional Neural Networks can take multidimensional arrays as inputs, therefore, an image or a stack of images are ideal for this architecture, although if any set of features can be expressed as such, they can also use CNN [4].

C. Spectrograms

An easy way of converting a speech signal into a multi-dimensional array is to convert it into some form of image. Majority of studies that opt for this method use a spectrogram. Spectrograms have been widely used as a speech representation for the amount of information they retain while being low dimension. A model trained by Badshah et al. [10] on Emo-DB used a deep CNN to extract discriminative features from spectrograms. The model had an accuracy of 84.3% when predicting 7 emotions. A different study by Wani et al. [9] is very similar as it used spectrogram inputs to a stride-based CNN. This study managed to set new benchmarks for the two datasets it used.

There are limitations as to how well a spectrogram can represent a speech signal. Spectrograms, being a product of the Fourier transform, does not represent abrupt changes efficiently as it assumes that a signal within a frame is stationary. Therefore, it is only possible to know which frequencies exist at what time interval, not instance. Furthermore, a fixed length of the window means time and frequency resolution are fixed for the entire length of the signal. The wavelet transform solves these problems by involving multi-resolution analysis [11]. A wavelet transform decomposes a function or a signal into a set of wavelets, wave-like oscillations localized in time. It is a very similar process to the Fourier transform, except that instead of convolving with sine waves, the wavelet transform convolves the signal with wavelets at varying scales and positions.

D. Scalograms

A continuous wavelet transform produces a scalogram image. Much like a spectrogram image, a scalogram is able to compress and retain a lot of information regarding the speech signal while staying low dimension. Applications of scalograms to SER in studies have been less common as compared to spectrograms. Some of the few examples include a study by Powroznik et al. [12] used fuzzy neural networks to identify speech emotion in 4 different languages with accuracy ranging from 62% to 94%. Another study [13] used scalogram inputs for SER by utilizing an attention-based BLSTM which achieved an accuracy of as high as 92% for certain emotions.

III. Methodology

In this chapter, we discuss the pre-processing of the speech signal, go over their conversion into their respective representations, and talk about the training of various neural network models. The flowchart in Fig. 1 can be used as a reference for this objective.

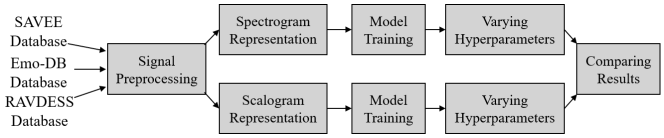

Fig. 1. Flowchart of the general methodology. Key milestones and processes are highlighted in gray.

A. Signal Pre-processing

To reduce speaker variations, the signals were normalized to have zero mean and unit variance [6]. This was done by taking the average and subtracting it from the signal, and then dividing the signal by its standard deviation. The resulting signal was then adjusted to have a length of 4 seconds as the databases' speech files had lengths varying from 2-4.5 seconds. This step was done so that every speech signal would have a uniform length. A paper by Izard [1] supports this process by stating that human emotion manifests itself in speech in bursts of 2-4 seconds. If an utterance had a length shorter than 4 seconds, it was zero-padded to achieve that length. For cases where the original length exceeded 4 seconds, the excess part at the end of the speech was removed to shorten it to the desired length. This portion of the methodology was conducted through MATLAB. The most noticeable change after the pre-processing was the change of scale and magnitude as compared in Fig 2. Other than that, there is no noticeable difference in the form and shape of the 2 signals.

B. Scalogram Generation

The scalogram of the processed signal was obtained through the use of the Continuous Wavelet Transform (CWT). The CWT of a signal, f(t), was computed using the equation below:

$$W(a,b) = \frac{1}{|a|^{\frac{1}{2}}} \int_{-\infty}^{+\infty} f(t)\bar{\psi}\left(\frac{t-b}{a}\right) dt \qquad (1)$$

$a \in R+$ is the scale value while $b \in R+$ is the transitional value. The ψ(t) is a function that remains continuous in both the time and the frequency domain, and is also called the mother wavelet; $\bar{\psi}$(t) is its complex conjugate. Daughter wavelets are produced by the mother wavelet by varying

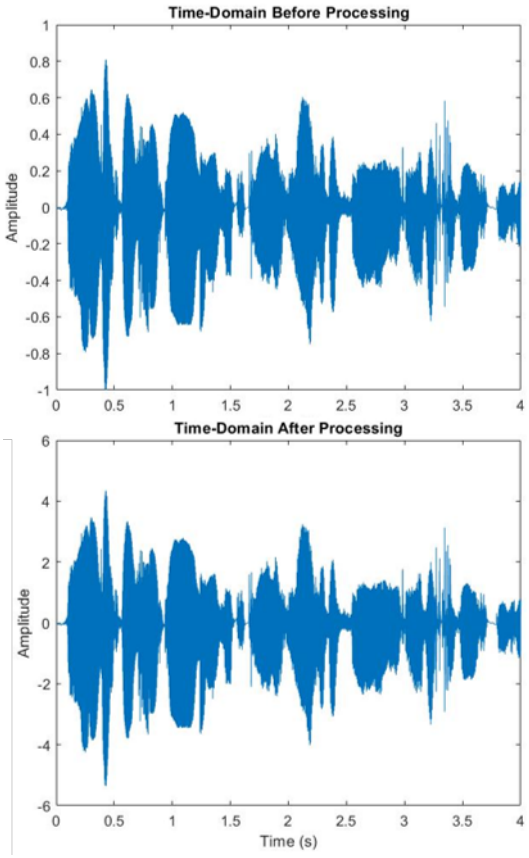

Fig. 2. (top) Time-domain of a speech signal before zero mean and unit variance. (bottom) Time-domain of speech signal after zero mean and unit variance.

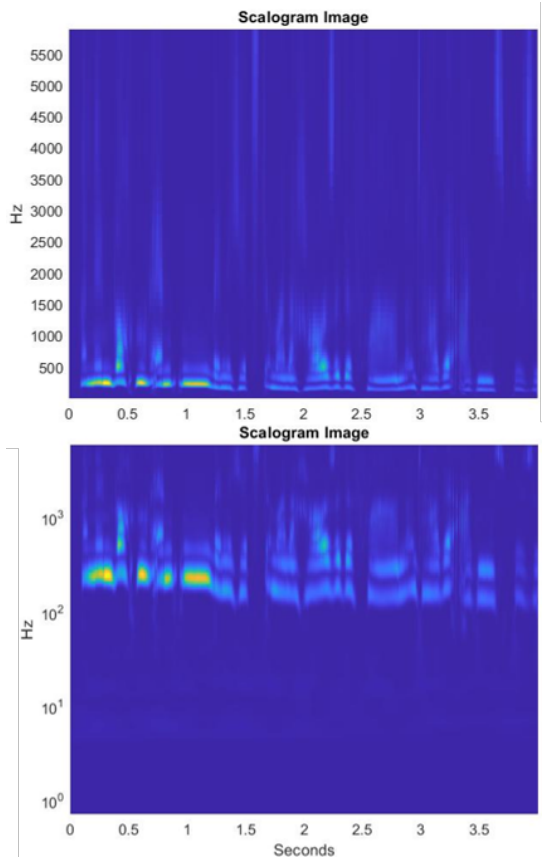

Fig. 3. (top) Scalogram with linear scaling. (bottom) Scalogram with log scaling.

different pairs of values of (a,b). For this purpose, the Morlet wavelet served as the mother wavelet as it is commonly used in applications for speech processing [14]. The Morlet wavelet is characterized by Equation 2.

$$\psi(\mathrm{t}) = \exp\left(-\frac{t^2}{2\sigma^2}\right)\exp(i\xi \mathrm{t}) \tag{2}$$

The typical values for σ(the width of the Gaussian), and ξ(specification of time-frequency trade-off) are 1 and 5 respectively. These were the same settings used for this experiment. The absolute values of the CWT coefficients were plotted to create a scalogram. Scalograms usually come with a default cone of influence in the image. As this will interfere with the CNN during input, this setting was removed from the scalogram. Furthermore, the scaling was changed to a log setting as illustrated in Fig 3. The scalogram image was further resized into a 256 x 256 x 3 image. This portion of the methodology was achieved through MATLAB.

C. Spectrogram Generation

The spectrogram of the processed signal was obtained through the use of Discrete Fourier Transform (DFT). The DFT can be obtained through the following equation:

$$X_n(\omega) = \sum_{m=-\infty}^{+\infty} x(m)w(n-m)e^{-j\omega \mathrm{m}} \tag{3}$$

The frequency component is ω while the windowing function is w. $X_n(\omega)$ can further be expressed as:

$$X_n(\omega) = |X_n(\omega)|\, e^{\mathrm{j}\arg[X_n(\omega)]} \tag{4}$$

The magnitude component is $|X_n(\omega)|$ and it was used to plot the spectrogram. MATLAB was used to create the spectrogram images. The sampling frequency that was used was dependent on the sampling frequency used for the speech signals in their respective datasets. The sampling rates are 16 kHz, 48 kHz, and 44.1 kHz for Emo-DB, RAVDESS, and SAVEE respectively. Meanwhile, Hamming, 50 dB, and 50% were used for the windowing method, maximum amplitude value, and overlap respectively. These settings were based on the settings used in a study by Ozseven [15] which also used spectrograms for SER. The spectrogram was further resized into a 256 x 256 x 3 image. This portion of the methodology was achieved through MATLAB.

D. Varying Hyperparameters and Retraining

To gather data, 8 base models based on the input type and database of origin were varied (See Table I for list of Base

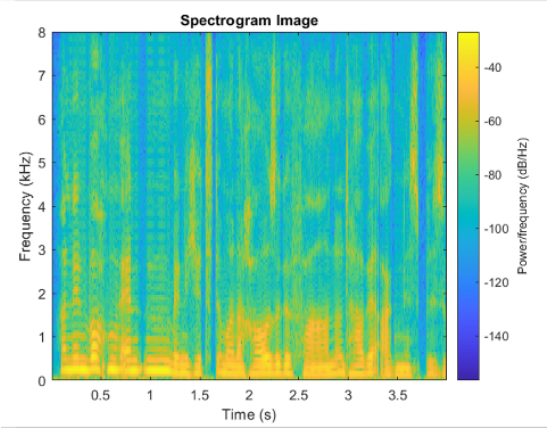

Fig. 4. Sample colored spectrogram image.

TABLE I
LIST OF BASE CNN MODELS.

Model	Database	Input Type
Model 1	Emo-DB	Spectrogram
Model 2	Emo-DB	Scalogram
Model 3	SAVEE	Spectrogram
Model 4	SAVEE	Scalogram
Model 5	RAVDESS	Spectrogram
Model 6	RAVDESS	Scalogram
Model 7	Mixed	Spectrogram
Model 8	Mixed	Scalogram

Models). The hyperparameters varied were the dropout rate and the hidden neuron number. The hidden layer neuron count was either reduced by 25%, maintained, or increased by 25%. The dropout rate for all layers was also either set to 20%, 35%, or remained at 50%. The hyperparameters add up to a combination of 9 different pairs for each base model. The base CNN model architecture that was varied can be seen in Fig. 5.

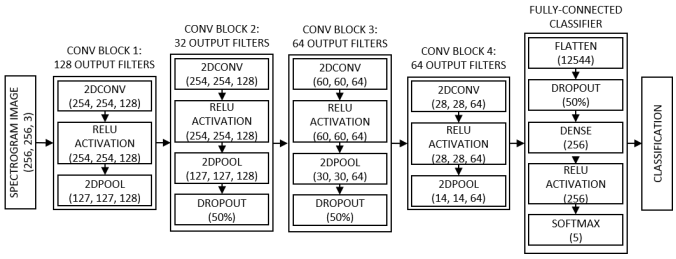

Fig. 5. Third CNN architecture derived from VGG16 (Model 3).

IV. RESULTS AND DISCUSSION

In this chapter, we present our results as well as an analysis and interpretation of said results. We conducted an analysis of variance to draw conclusions from. To supplement the ANOVA test, an estimation of difference between two population means was also applied.

TABLE II
SCALOGRAM WEIGHTED F1-SCORE RESULTS SUMMARY FROM 108 TRAINED MODELS.

SCALOGRAM RESULTS (weighted f1-score)				
	Berlin Scalogram	SAVEE Scalogram	RAVDESS Scalogram	Mixed Scalogram
(312, 0.20)	0.69	0.57	0.62	0.61
(312, 0.35)	0.69	0.57	0.62	0.62
(312, 0.50)	0.58	0.55	0.61	0.63
(416, 0.20)	0.67	0.56	0.64	0.59
(416, 0.35)	0.66	0.54	0.66	0.60
(416, 0.50)	0.66	0.54	0.68	0.62
(520, 0.20)	0.70	0.49	0.67	0.60
(520, 0.35)	0.67	0.55	0.64	0.60
(520, 0.50)	0.65	0.53	0.63	0.61

TABLE III
SPECTROGRAM WEIGHTED F1-SCORE RESULTS SUMMARY FROM 108 TRAINED MODELS.

SPECTROGRAM RESULTS (weighted f1-score)				
	Berlin Spectro	SAVEE Spectro	RAVDESS Spectro	Mixed Spectro
(312, 0.20)	0.67	0.58	0.70	0.59
(312, 0.35)	0.69	0.59	0.64	0.58
(312, 0.50)	0.65	0.48	0.65	0.55
(416, 0.20)	0.61	0.53	0.67	0.59
(416, 0.35)	0.69	0.53	0.62	0.58
(416, 0.50)	0.66	0.52	0.61	0.58
(520, 0.20)	0.55	0.50	0.71	0.56
(520, 0.35)	0.65	0.54	0.67	0.59
(520, 0.50)	0.57	0.50	0.64	0.57

A. Analysis of Variance

Accuracy (correct predictions over total number of predictions) is not always the most reliable performance metric in classification problems [16]. It is important to have a metric that considers both the precision and the recall of the models. Precision is the sum of true positives across all classes divided by the sum of true and false positives in all classes, while recall is the sum of true positives across all classes divided by the sum of false negatives and true positives across all classes. The F1-score combines these two through equation 5. The F1-score works well in summarizing the performance in classification problems, especially when dealing with unbalanced datasets.

$$F1 = \frac{2x \text{ Precision } x \text{ Recall}}{\text{Precision } + \text{ Recall}} \tag{5}$$

Table II and III show the results of the trained models in terms of weighted f1-score. The values in these tables were also put through an ANOVA test and the results are shown in Table IV.

The P-value, 0.293350 is greater than the alpha value of 0.05. Furthermore, the F-value is less than the F critical value. These two facts allow us to conclude that there is not enough evidence to reject the null hypothesis, which is that there is no difference between the mean weighted F1-scores of the scalogram-based models and the spectrogram-based models.

TABLE IV
SINGLE-FACTOR ANOVA TEST RESULTS AND SUMMARY AT 95% CONFIDENCE LEVEL FOR WEIGHTED F1-SCORE.

ANOVA: Single Factor

SUMMARY	**Groups**	**Count**	**Sum**	**Average**	**Variance**		
	Scalogram	36	22.12	0.614444	0.002660		
	Spectrogram	36	21.61	0.600278	0.003786		
ANOVA	**Source of Variation**	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
	Between Groups	0.003613	1	0.003613	1.120969	0.293350	3.977779
	Within Groups	0.225586	70	0.003223			
	Total	0.229199	71				

B. Estimation of the Difference between Means

To supplement the results of this ANOVA test for the weighted f1-scores, we find the estimated difference between the two population means at the same confidence level of the ANOVA test, 95%. Plugging in the values in Equation 6 where $\bar{X}$ is the sample mean of the scalogram-based models and $\bar{Y}$ is the sample mean of the spectrogram-based models, we get the range (-0.012059, 0.040392). This range contains the value 0, therefore, we can conclude that at a 95% interval, the mean weighted f1-score of the scalogram-based models is not generally greater than the mean of the spectrogram-based models. This supports the conclusion from the ANOVA test done on the weighted f1-scores which says that the difference between the two population means is not statistically significant.

$$\left((\bar{X}-\bar{Y}) - Z_{\frac{\alpha}{2}}\sqrt{\frac{s_X^2}{n_1}+\frac{s_Y^2}{n_2}}, (\bar{X}-\bar{Y}) + Z_{\frac{\alpha}{2}}\sqrt{\frac{s_X^2}{n_1}+\frac{s_Y^2}{n_2}}\right) \quad (6)$$

C. Computational Complexity

For this study, both the spectrogram conversions and scalogram conversions were conducted in MATLAB through the use of the spectrogram() and cwt() functions respectively. Anecdotally, the spectrogram conversion ran significantly faster than that of the scalogram conversion although some literature [17] would suggest otherwise. However, it is important to note that these conversions were optimized under the hood by MATLAB; the exact information regarding these optimizations is not readily available and further investigation is required.

MATLAB constructs a spectrogram by taking a series of FFTs and overlapping them [18]. An FFT is a more efficient implementation of the DFT that runs with a computational complexity of $O(nlog_2 n)$ while the latter runs with $O(n^2)$ [17]. As for the cwt() function, it is described on the website as a discretized version of the CWT that takes much larger computational resources than the DWT [19]. Studies have presented methods for implementing the CWT at a time complexity of $O(n)$ which is asymptotically faster than the FFT [20, 21] but whether this is how MATLAB applies the cwt function does not seem to be the case. Table V documents the amount of time it takes to convert a signal into a scalogram and a spectrogram in MATLAB. The *timeit* function was used to record the time and the codes run were as bare as possible – only reading the audio file and converting it immediately. It can be observed how the conversions take roughly the same amount of time for signals with lengths of 1 second. However, the scalogram conversion scaled horribly with respect to the signal length. The change from 1 second to 50 seconds saw an 800 times increase in conversion time for scalogram conversions while spectrogram conversions only experienced a 1.5 times increase.

TABLE V
TIME TO CONVERT SIGNAL TO SCALOGRAM VS SPECTROGRAM

Length of sound file (seconds)	**Time to convert to scalogram (seconds)**	**Time to convert to spectrogram (seconds)**
1	0.3778	0.370
4	1.0779	0.370
10	3.2178	0.381
20	6.4289	0.395
50	308.11	0.545

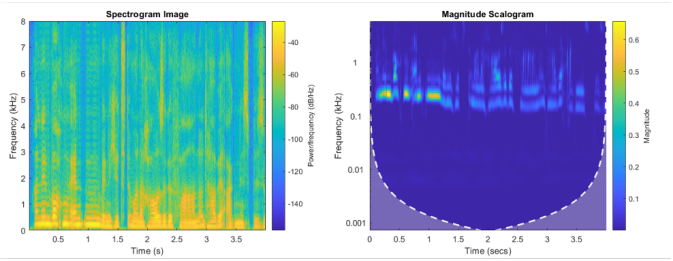

Fig. 6. Sample scalogram and spectrogram images.

The size of the input contributes greatly to the computational complexity of training and testing models. It can be observed in Fig. 6 that the information contained in the scalogram images are concentrated in the upper portion while the bottom half remains largely empty. The difference is great when comparing to spectrograms with information covering the entirety of the image. This means that there is a unique potential for the size of the scalogram input to be reduced by half, while still maintaining the same performance as practically no information is lost. This may give scalograms an edge as inputs as it will lighten the load for computations when training and testing a model.

Overall, spectrograms seem to have the advantage of being far faster to convert in MATLAB while scalograms have the

potential to ease computational burdens as an input to neural networks.

V. Conclusion

In this paper, we sought to determine the capabilities of a scalogram representation input in SER problems using CNN. We hypothesized that based on the scalogram representation's performance in other classification problems, it had the potential to surpass spectrogram representations for SER. To this end, we have processed and converted a total of 1,615 speech utterances from 3 databases (Emo-DB, SAVEE, RAVDESS) into uniform spectrogram and scalogram images which were used as inputs to variations of a general image classification CNN model. The models were assessed on how well they identified 5 emotions: Anger, Disgust, Happiness, Fear, and Neutral.

After training 108 models for each representation and comparing the results of 36 from each, the ANOVA tests have shown that spectrograms and scalograms have statistically the same performance. Future work is needed to confirm whether the same conclusions hold for more complex architectures and to find cases in which one representation is more applicable than the other.

Acknowledgment

We would like to thank the UP Digital Signal Processing laboratory for the support and resources they provided. [1] would also like to acknowledge DOST-SEI for providing me with the funds to accomplish this project. [2] would like to thank Dr. Andrew Hines, QxLab, ADAPT-Centre, and Insight for their guidance and support.

This research was conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at University College Dublin. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

[1] C. Izard, *Human Emotions*, en. Springer Science and Business Media, 2013.

[2] G. Lu, L. Yuan, and W. Yang, "Speech emotion recognition based on long-term and short-term memory and convolutional neural network," en, *J. Nanjing Univ. Posts Telecommun*, vol. 38, no. 5, 63–69, 2018.

[3] W. Lim, D. Jang, and T. Lee, *Speech emotion recognition using convolutional and recurrent neural networks*, en, 2016.

[4] D. Issa, M. Demirci, and A. Yazici, "Speech emotion recognition with deep convolutional neural networks," en, *Biomedical Signal Processing and Control*, vol. 59, May 2020. DOI: 10.1016/j.bspc.2020.101894..

[5] C. Anagnostopoulos, T. Iliou, and I. Giannoukos, "Features and classifiers for emotion recognition from speech: A survey from 2000 to 2011," en, *Artificial Intelligence Review*, vol. 43, no. 2, 155–177, Feb. 2015. DOI: 10.1007/s10462-012-9368-5..

[6] H. Fayek, M. Lech, and L. Cavedon, "Evaluating deep learning architectures for speech emotion recognition," en, *Neural Networks*, vol. 92, 60–68, Aug. 2017. DOI: 10.1016/j.neunet.2017.02.013..

[7] L. Sun, B. Zou, S. Fu, J. Chen, and F. Wang, "Speech emotion recognition based on dnn-decision tree svm model," en, *Speech Communication*, vol. 115, 29–37, Dec. 2019. DOI: 10.1016/j.specom.2019.10.004..

[8] A. Qayyum, A. Arefeen, and C. Shahnaz, "Convolutional neural network (cnn) based speech-emotion recognition," en, in *2019 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON*, 2019.

[9] T. Wani, T. Gunawan, S. Qadri, H. Mansor, F. Arifin, and Y. Ahmad, "Stride based convolutional neural network for speech emotion recognition," en, in *2021 IEEE 7th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA*, Aug. 2021, 41–46. DOI: 10.1109/ICSIMA50015.2021.9526320..

[10] A. Badshah, J. Ahmad, N. Rahim, and S. Baik, "Speech emotion recognition from spectrograms with deep convolutional neural network," en, in *2017 International Conference on Platform Technology and Service (PlatCon*, 2017.

[11] A. Akansu and R. Haddad, "Time-frequency representations," fr, *Multiresolution Signal Decomposition*, 331–390, 2001.

[12] P. Powroznik, P. Wojcicki, and S. Przylucki, "Scalogram as a representation of emotional speech," en, *IEEE Access*, vol. 9, 154044–154057, 2021. DOI: 10.1109/ACCESS.2021.3127581..

[13] K. Aghajani and I. P. Afrakoti, "Speech emotion recognition using scalogram based deep structure," en, *International Journal of Engineering, Transactions B: Applications*, vol. 33, no. 2, 285–292, Feb. 2020. DOI: 10.5829/IJE.2020.33.02B.13..

[14] K. Aghajani and I. P. Afrakoti, "Speech emotion recognition using scalogram based deep structure," en, *International Journal of Engineering, Transactions B: Applications*, vol. 33, no. 2, 285–292, Feb. 2020. DOI: 10.5829/IJE.2020.33.02B.13..

[15] T. Özseven, "Investigation of the effect of spectrogram images and different texture analysis methods on speech emotion recognition," en, *Applied Acoustics*, vol. 142, 70–77, Dec. 2018. DOI: 10.1016/j.apacoust.2018.08.003..

[16] A. Gupta and A. Yilmaz, "Social network inference in videos," en, in *Academic Press Library in Signal Processing: Image and Video Processing and Analysis and Computer Vision*, vol. 6, Elsevier, 2018, 395–424. DOI: 10.1016/B978-0-12-811889-4.00011-7..

[17] K. Wirsing, "Chapter: Time frequency analysis of wavelet and fourier transform," en, *IntechOpen*, Nov. 18, 2020.

[18] M.A.T.L.A.B., *Spectrogram using short-time fourier transform*, en.

[19] M.A.T.L.A.B., *Continuous and Discrete Wavelet Transforms*, pt. MATLAB and Simulink".

[20] A. Muñoz, R. Ertlã, and M. Unser, *Continuous wavelet transform with arbitrary scales and o(n) complexity*, en, Online]. Available: www.elsevier.com/locate/sigpro, 2002.

[21] Sadowsky, *The continuous wavelet transform: A tool for signal investigation and understanding*, no.