Motion sensor data anonymization by time-frequency filtering

Noëlie Debs, Théo Jourdan, Ali Moukadem, Antoine Boutet, Carole Frindel

Motion sensor data anonymization by time-frequency filtering

Noëlie Debs*, Théo Jourdan*‡, Ali Moukadem†, Antoine Boutet‡, Carole Frindel*,
*Univ Lyon, INSA Lyon, CREATIS, Lyon, France
noelie.debs@creatis.insa-lyon.fr, carole.frindel@creatis.insa-lyon.fr
†Université Haute-Alsace, IRIMAS, Mulhouse, France
ali.moukadem@uha.fr
‡Univ Lyon, INSA Lyon, Inria, CITI, Lyon, France
theo.jourdan@insa-lyon.fr, antoine.boutet@insa-lyon.fr

***Abstract*—Recent advances in wireless actimetry sensors allow recognizing human real-time activities with mobile devices. Although the analysis of data generated by these devices can have many benefits for healthcare, these data also contains private information about users without their awareness and may even cause their re-identification. In this paper, we propose a privacy-preserving framework for activity recognition. The method consists of a two-step process. First, acceleration signals are encoded in the time-frequency domain by three different linear transforms. Second, we propose a method to anonymize the acceleration signals by filtering in the time-frequency domain. Finally, we evaluate our approach for the three different linear transforms with a neural network classifier by comparing the performances for activity versus identity recognition. We extensively study the validity of our framework with a reference dataset: results show an accurate activity recognition (85%) while limiting the re-identifation rate (32%). This represents a large utility improvement (19%) against a slight privacy decrease (10%) compared to state-of-the-art baseline.**

***Index Terms*—Activity Recognition, Privacy, Time-Frequency, Classification, Convolutional Neural Networks**

I. Introduction

The emergence of Internet of Things (IoT) devices have paved the way for personal monitoring. These devices record electronic measurements from a variety of sensors (accelerometer, gyroscope and magnetometer) and send the person data to an application server to be analyzed. This analysis implies advanced signal processing and machine learning algorithms to provide a variety of services such as number of steps, burned calories, traveled distance and sleep monitoring [1]. However, the data captured by these sensors can also contains private information about users without their awareness where highly sensitive information can be inferred such as the user's health status [2], [3] or even the user identity [4]. Yet, the complex workflow of collected data multiplies the security and privacy risks, including the data collection and transmission [5], as well as the processing and the storage [6].

In this work, we explore a new method for anonymizing motion sensor data by obfuscating biometric content in order to avoid the re-identification of the persons, while preserving the remainder of the activity pattern. Achieving this balance between data utility (i.e., the activity recognition) and data privacy (i.e., avoid users re-identification) is an important objective to send secure and reliable data through mobile devices and to strengthen end-user confidence. This utility and privacy trade-off has been analyzed in different contributions for motion sensor data. Privacy-preserving mechanisms for time-series data can be implemented through different mechanisms. Perturbation approaches hide sensitive patterns by adding a crafted noise to each time-window of the time-series [7], [8]. Synthesis approaches generate data that maintain some required statistics of the original data without information that can be used for re-identification [9], [10]. In a similar way, transformation approaches can reduce the amount of sensitive information in the data by reconstruction [11]. Finally, filtering approaches can be used to remove unwanted components only in time-series parts that include sensitive information. For example, [12] suppresses location time-series when users are in a sensitive place, such as a hospital. Interestingly, [13] shows that features in temporal domain are useful to discriminate user activity while features in frequency domain lead to discriminate the user identity.

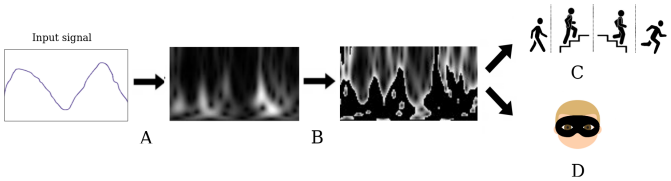

Fig. 1. Overview of the proposed pipeline, divided in 4 steps: **A.** Signal transformation into a time-frequency (TF) image, **B.** Anonymization method based on image filtering, **C.** Activity recognition and **D.** User identification

As information patterns are contained in the temporal and in the frequency domain, our approach relies on time-frequency (TF) representation. In addition, to improve privacy, sensitive information are removed from this TF representation. More precisely, since motion sensors respond to both frequency and intensity of movements and their outputs are non-stationary signals, Fourier transform that provides frequency components of the whole signal is not sufficient to describe the signal properly. In this context, we propose to use a TF encoding of the signal to learn to recognize activity on one side and identity on the other. This learning step is based on a neural network

which allows good learning performance and the automatic selection of descriptors of interest in the encoding space. Finally, we propose to filter sensitive identity information in this representation by canceling the highest value coefficients which correspond to the person's activity rate. The validity of our framework is extensively demonstrated on a public dataset and related to a state-of-the-art baseline.

The article is organized in the following way. Time-frequency transforms and CNN classifier are detailed in Section II. Then tools used for the evaluation are described in Section III. Section IV presents the results before to conclude in Section V.

II. Material and Method

A. Dataset

We used the public dataset Motion-Sense to assess the performance of our approach. It includes data sensed from 3-axis motion sensors at a constant frequency of 50 Hz collected with an iPhone 6s kept in the participant's front pocket [14]. Overall, a total of 24 participants have performed six activities during 15 trials in the same environment and conditions. The considered activities are going downstairs, going upstairs, walking, jogging, sitting and standing. To enable both classifications over time (i.e., the activity and the identity), the raw motion data are split in sliding windows, where each sliding window is a sample of a single activity. Knowing that in average the cadence range of walking is not less than 1.5 steps per second [15], the window length is chosen to be 2.5 seconds with an overlap of 50 %.

For this study, we focused on the four dynamic activities (going upstairs, going downstairs, walking and jogging): they are the most difficult to analyze and their complex frequency content is adapted to TF representation. Also, we only focused on the user acceleration signal, which is adapted for dynamic activities [16].

B. Time-frequency domain

The time-frequency domain tries to overcome the limitations of the classical Fourier transform, which only provides frequency content without any time information. Indeed, in the context of non-stationary signals, we need to know the frequency evolution of the signal components as a function of time. There are several approaches to project the signal in the TF domain. In this paper we will focus on 3 different linear transforms: The STFT [17], the Stockwell transform [18] and the optimized Stockwell transform [19].

1) The Short-Time Fourier Transform: Given a signal $x(t)$, its Short-Time Fourier Transform (STFT) can be given as:

$$Sf_x(t,f) = \int_{-\infty}^{+\infty} x(\tau) w^*(\tau - t) e^{-2j\pi f \tau} d\tau, \qquad (1)$$

where $w(t)$ is the analysis window which has a fixed width (mono-resolution analysis). The window $w(t)$ is chosen in this paper as a Gaussian function with a standard deviation $\sigma = 0.05$.

2) The Stockwell transform: The Stockwell transform (S-transform) can be considered as a hybrid between the STFT and the Continuous Wavelet Transform (CWT) [20]. It preserves a direct relation with the Fourier's kernel as the STFT, while performing a multiresolution analysis as the CWT. The S-transform of signal $x(t)$ can be expressed as follows:

$$S_x(t,f) = \int_{-\infty}^{+\infty} x(\tau) w^*(\tau - t, f) e^{-2\pi j f \tau} d\tau, \qquad (2)$$

where $w(t,f)$ is the analysis window which is a Gaussian function of two variables: time and frequency. It can be given as:

$$w(t,f) = \frac{1}{\sigma(f)\sqrt{2\pi}} e^{\frac{-t^2}{2\sigma(f)^2}} \qquad (3)$$

The standard deviation $\sigma(f)$ is inversely proportional to the frequency:

$$\sigma(f) = \frac{1}{|f|}, \qquad (4)$$

to promote temporal resolution for low frequencies and frequency resolution for high frequencies.

3) Optimized S-transform: To better adapt the analysis window to the nature of the signal being analyzed, many authors tried to optimize the S-transform representation by introducing new parameters on the Gaussian window [21], [22]. Among these proposals, a generalized Gaussian window controlled by a set of parameters was proposed in [19] as follows :

$$w^{p_i}(t,f) = \frac{|f|^r}{(mf^p + k)\sqrt{2\pi}} e^{\frac{-(\tau-t)^2 f^{2r}}{2(mf^p+k)^2}}. \qquad (5)$$

The idea is to choose the set of parameters $p_i \in \{r, m, p, k\}$ that maximizes an energy concentration function. This energy concentration can be measured by several approaches. The concentration measurements (CM) used in this paper are given as follows [23]:

$$CM(p_i) = \frac{1}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left|\overline{S_x^{p_i}(t,f)}\right| dtdf}, \qquad (6)$$

with $\overline{S_x^{p_i}(t,f)}$ a normalization of $S_x^{p_i}(t,f)$ [22]:

$$\overline{S_x^{p_i}(t,f)} = \frac{S_x^{p_i}(t,f)}{\sqrt{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |S_x^{p_i}(t,f)|^2 dtdf}}. \qquad (7)$$

The parameters p_i which maximize $CM(p_i)$ are chosen to compute the optimized S-transform. In this study, the optimization is carried out on the whole signal and the optimal parameters were calculated for a sample of the signals in the dataset. This allows to observe the trend of the variation of these parameters particularly for the various activities where no significant variation is observed. In our case, the parameters are fixed as follows: $r = 0.7$, $m = 0$, $p = 0$ and $k = 0.4$.

C. Identity filtering

TF images generated from the different TF transformations have a size of 62 and 128 pixels in spectral (frequency voices) and temporal domains respectively. As depicted on Figure 2, TF images for walking activity present different patterns from TF images for running activity, and can be discriminated in terms of texture: especially the number of vertical salient peaks. On the other hand, two different users can present differences in the contrast of their TF image as shown in Figure 2, where peaks of user #8 are more contrasted than those of user #15. These observations stress the interest of filtering high coefficients to remove user information.

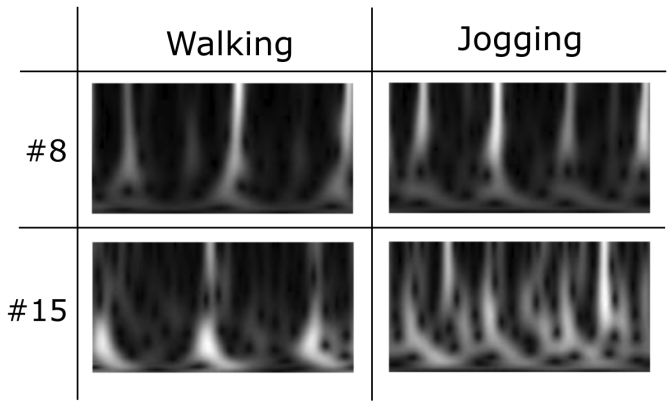

Fig. 2. Representation of optimized S-transform for 2 different users (#8 and #15) for two different activities (walking and jogging).

In agreement with these observations, identity filtering consists in setting different percentages x of the total TF image coefficients (sorted in descending order) to zero : x ranging from 10% to 90% with a step of 10%. This method allows us, in agreement with [13], to ensure that information relevant to re-identification is removed first.

D. CNN classifier

We propose two distinct convolutional neural networks as classifier to assess the performance of our framework: one that classifies signals into 4 classes (corresponding to the 4 activities) called CNN activity, and another that classifies the same signals into 24 classes (corresponding to the 24 subjects in the study) called CNN identity. These two models have the same architecture, but are trained separately.

1) CNN Architecture: Multi-inputs image classification based on CNNs can be addressed using early fusion strategy, where all input images are combined at the beginning of the network. This fusion strategy present low computational complexity and is an easy implementation [24]. However, it has been shown in other contexts that the late fusion better accounts the complexity of each input and outperforms the early fusion [25], [26]. The late fusion strategy consists in processing each input image independently on distinct convolutional branches, and merging features at a higher level.

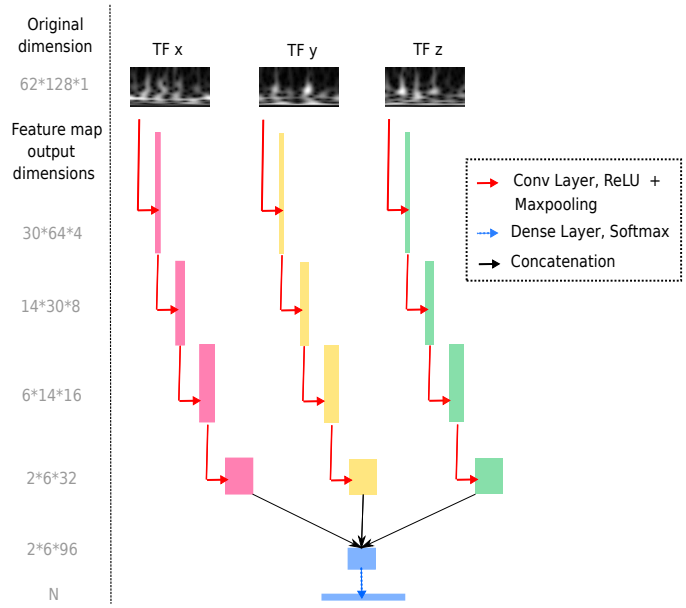

Fig. 3. Overview of the proposed CNN architecture. The network takes three TF images (TF_x, TF_y, TF_z) as input. Each input image is processed independently on 3 separate branches. Pink, yellow and green feature maps result from 2D-convolutions and maxpooling. The output of the 3 branches are then concatenated, and passed through a hidden layer of N nodes, with $N = 4$ for CNN activity and $N = 24$ for CNN identity.

In our case, each signal window was defined by three different TF images, standing for the acceleration along the x, y and z axes (TF_x, TF_y, TF_z, respectively). The detailed architecture is depicted in Figure 3.

2) CNN implementation: For both CNN activity and CNN identity, signals in the dataset have been split according to trials: 90% of signals from trials 1,2,3,4,7,8,9 were used as training set, and the remaining 10% as validation set. Signals from trials 11,12,15,16 were used as testing set. For both CNNs, we used a categorical cross-entropy loss function that produced weights to equally penalize under or over-represented classes in the training set. The optimizer was set with Adam, the batch size was set to 128, and the number of epochs was set to 150 but was regulated by early stopping. The total number of weights to train was 23,044 for CNN activity and 46,104 for CNN identity.

III. Evaluation

A. Classification metric

To assess the classification performances of the two CNNs (activity and identity), we computed for each class the accuracy metric acc, defined as:

$$acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{8}$$

where TP stands for true positives, TN for true negatives, FN for false negatives and FP for false positives.

We called the result respectively Activity acc (i.e. data utility) and Identity acc (i.e. privacy) when it is applied to

the activity recognition and to the user identity. The given accuracies systematically corresponds to the results averaged over ten experiments.

B. State-of-the-art baseline

To compare the performance obtained from the different TF representations, we proposed a baseline based on the Fourier transform of the acceleration signal. In this baseline, just like our filtering approach, we filtered different percentages x of the transform coefficients in descending order (x ranging from 10% to 90% with a step of 10%). Once Fourier transform was filtered, as did [13], the signal was classified into activity and identity classes based on frequency domain features using a Random Forest classifier.

C. Optimal representation metric

We defined the best TF representation as the one that maximizes the Area Under the utility-privacy Curve (AUC) using the trapezoidal rule. The AUC is an effective and combined measure of utility and privacy that describes the inherent validity of the anonymization approach. The AUC used here was bounded in x between the minimum and maximum performance in identity. As these bounds changed between the different TF representations, we normalized the AUC by the size of the rectangle relating to the bounds.

D. Optimal filter

The optimal filter was defined as the filter minimizing the Euclidean distance between a point from the utility-privacy curve and the upper left corner of normalized area – that corresponds to the intersection of the upper edge of the Figure 4 (maximal performance of 100% in activity) and the minimum bound in performance for the identity. This optimal filter guarantees a good activity recognition while limiting identity identification.

IV. Results and Discussion

A. Interest of the time-frequency representation

Figure 4 shows the filtering effect on activity and identity recognition for the three different TF representations and the baseline. It appears that working only in the frequency domain (cross markers in Figure 4) leads quickly to a significant loss of activity recognition. This loss in data utility is explained by the fact that Fourier transform is not able to correctly analyze the non-stationary nature of signals, where frequency content changes over time especially in the activity pattern. On the other hand, TF representations seem to be able to deal with the non-stationarity of the signals and therefore allow a better trade-off between activity recognition and user identification. The more coefficients are filtered from the TF images, the worse identification performance is. Conversely, filtering has much less impact on activity recognition performance: whether no filtering is applied or that 70% of the image is filtered, activity accuracy seems to be stable between 80% and 90%. This trend is observed whatever the TF representation (STFT, S-transform or optimized S-transform respectively round, square

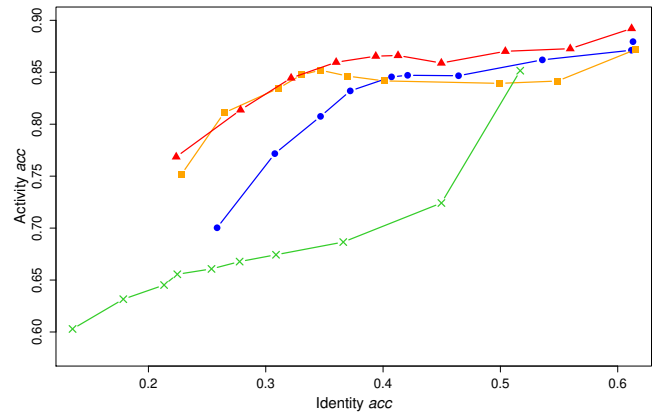

Fig. 4. Activity accuracy according to identity accuracy for different representations: the Fourier transform (cross markers in green), the STFT (round markers in blue), the S-transform (square markers in orange) and the optimized S-transform (triangle markers in red). Each point corresponds to an average classification result over 10 experiments. The upper left corner represents the ideal trade-off between utility and privacy. For each curve, the high performance points in activity and in identity correspond to cases without filtering while the others (as one tends to the left of the graph) correspond to filtering cases with a step of 10%.

and triangle markers in Figure 4). These results demonstrate that high coefficients in the TF images carry specifically the person's activity rate and hence the identity information. Activity, on the other hand, seems less specific to a range of coefficients and corresponds more to the general texture observed in the TF images.

B. Optimal representation

Table I summarizes the normalized AUC for each curve in the Figure 4. Among the three TF representations, STFT has the lower AUC ($AUC = 0.83$), hence offering a worse trade-off between utility and privacy. This observation can be linked to the fact that this transform is mono-resolution and therefore provides a lower time-frequency resolution of the associated signal. On the other hand, the S-transform is multi-resolution which allows a better encoding of the analyzed signal and hence higher activity recognition ($AUC = 0.84$). Activity recognition is further improved when the S-transform of the acceleration signal is optimized according to an energy concentration criterion ($AUC = 0.85$). Differences between S-transform and optimized S-transform could possibly be even more marked if the S-transform optimization would have been done on individual sliding windows rather than the whole signal. Unsurprisingly, the better is the energy concentration in the time-frequency plane – which means a better tonal resolution of the TF image – the faster the neural network converges.

C. Optimal filter

Table II shows that the optimal filter that guarantees a good utility-privacy trade-off is 60% for the Fourier, STFT and S-transform representations and 70% for the optimized

TABLE I
NORMALIZED AREA UNDER THE UTILITY-PRIVACY CURVES (AUC) FOR EACH REPRESENTATION

Fourier	STFT	S-transform	Opti. S-transform
0.69	0.83	0.84	0.85

TABLE II
OPTIMAL FILTER IN % FOR EACH REPRESENTATION, AND THE ASSOCIATED PERFORMANCES IN % (ACTIVITY acc/IDENTITY acc)

Fourier	STFT	S-transform	Opti. S-transform
60 (66/22)	60 (83/37)	60 (85/33)	70 (85/32)

S-transform representation. These observations suggest that given the better tonal resolution of the optimized S-transform representation, it is possible to filter more coefficients of the TF image without losing too significantly in activity performance (as observed for the other TF representations in the Figure 4.

V. CONCLUSION

In this work, we presented a new proof of concept method for preserving individual privacy in motion sensor data. This method uses time-frequency representation of acceleration signals and filters the resulting TF images by setting the highest coefficients to zero before the machine-learning step. The evaluations demonstrated that our method successfully anonymized identity, and preserved a high activity recognition ratio by better encoding of the non-stationary aspect of the signals than the Fourier transform. More specifically, we determined that the optimized S-transform gives the best utility-privacy trade-off by filtering its TF coefficients at 70%. The proposed filtering privacy-preserving mechanism was intentionally simple, but show promising results. More advanced filtering methods [27] could be considered to improve performance, which will be the subject of future research. Moreover, other time-frequency transforms can be applied and compared with the results obtained in this paper.

REFERENCES

[1] J. M. Peake, G. Kerr and J. P. Sullivan, "A Critical Review of Consumer Wearables, Mobile Applications, and Equipment for Providing Biofeedback, Monitoring Stress, and Sleep in Physically Active Populations," Front Physiol., vol 9, 2018.

[2] A. Zhan, M. Chang, Y. Chen, and A. Terzis, "Accurate Caloric Expenditure of Bicyclists Using Cellphones," In 10th Conference on Embedded Network Sensor Systems, ACM, pp. 71–84, 2012.

[3] K. Plarre, A. Raij, S.M. Hossain, A. Ali, M. Nakajima, M. al'Absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smailagic, and L. Wittmers, "Continuous Inference of Psychological Stress from Sensory Measurements Collected in the Natural Environment," In 10th International Conference on Information Processing in Sensor Networks, ACM/IEEE, pp. 97-108, 2011.

[4] N. Neverova, C. Wolf, G. Lacey, L. Fridman, D. Chandra, B. Barbello and G. Taylor, "Learning Human Identity From Motion Patterns," Access, IEEE, vol 4, pp. 1810-1820, 2016.

[5] D. Wood, N. Apthorpe and N. Feamster, "Cleartext data transmissions in consumer iot medical devices, " In 2017 Workshop on Internet of Things Security and Privacy, ACM, pp. 7-12, 2017.

[6] M. Rushanan, A. D. Rubin, D. F. Kune and C. M. Swanson, "Sok: Security and privacy in implantable medical devices and body area networks," In Symposium on Security and Privacy, IEEE, pp. 524-539, 2014.

[7] Y. Amar, H. Haddadi and R. Mortier, "An information-theoretic approach to time-series data privacy," In 1st Workshop on Privacy by Design in Distributed Systems, ACM, p. 3, 2018.

[8] G. Acs and C. Castelluccia. "A Case Study: Privacy Preserving Release of Spatio-temporal Density in Paris." In 20th International conference on Knowledge discovery and data mining (SIGKDD), ACM, 2014.

[9] B. K. Beaulieu-Jones, Z. S. Wu, C. Williams, R. Lee, S. P. Bhavnani, J. B. Byrd and C. S. Greene, "Privacy-preserving generative deep neural networks support clinical data sharing, Circulation: Cardiovascular Quality and Outcomes," Circulation: Cardiovascular Quality and Outcomes, vol 12, 2019.

[10] G. Acs, L. Melis, C. Castelluccia and E. De Cristofaro, "Differentially private mixture of generative neural networks," Transactions on Knowledge and Data Engineering, IEEE, vol.31, pp. 1109–1121, 2019.

[11] M. Malekzadeh, R. G. Clegg, A. Cavallaro and H. Haddadi, "Mobile sensor data anonymization," In Proceedings of the International Conference on Internet of Things Design and Implementation, pp. 49-58, 2019.

[12] M. Götz, S. Nath and J. Gehrke, "MaskIt: Privately releasing user context streams for personalized mobile applications," In International Conference on Management of Data (SIGMOD), ACM, pp. 289–300, 2012.

[13] T. Jourdan, A. Boutet and C. Frindel, "Toward privacy in IoT mobile devices for activity recognition," In 15th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous), EAI, pp. 155-165, 2018.

[14] J. L. Reyes-Ortiz, "Smartphone-based human activity recognition," Springer, 2015.

[15] C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification," In 5th International Conference on Automatic Face Gesture Recognition, IEEE, pp. 372–377, 2002.

[16] N. Ravi, N. Dandekar, P. Mysore and M. Littman, "Activity Recognition from Accelerometer Data," AAAI, vol. 3, p. 1541-1546, 2005.

[17] D. Gabor, "Theory of Communication," Part 1, J. Inst. of Elect. Eng. Part III, Radio and Communication, vol 93, p. 429, 1946.

[18] R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: the S transform," Transactions on Signal Processing, IEEE, vol. 44, pp. 998-1001, 1996.

[19] A. Moukadem, Z. Bouguila, D.O Abdeslam and A. Dieterlen, "A new optimized Stockwell transform applied on synthetic and real non-stationary signals," Digital Signal Processing, vol. 46, pp. 226-238, 2015.

[20] S. Mallat, "A wavelet tour of signal processing 2nd Edition," Academic Press, 1999.

[21] S. Assous and B. Boashash, "Evaluation of the modified S-transform for time- frequency synchrony analysis and source localisation," European Association for Signal Processing (EURASIP) J. Adv. Signal Process, pp. 1–18, 2012.

[22] E. Sejdic, I. Djurovic and J. Jiang, "A window width optimized S-transform", European Association for Signal Processing (EURASIP) J. Adv. Signal Process, p. 59, 2007.

[23] L. Stankovic, "A measure of some time–frequency distributions concentration, Signal Process,", vol. 81, pp. 621–631, 2001.

[24] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. De Vries, M. Benders and I. Išgum, "Automatic segmentation of MR brain images with a convolutional neural network," Transactions on medical imaging, IEEE, vol. 35, pp. 1252-1261, 2016.

[25] N. Srivastava and R.R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," Advances in neural information processing systems, pp. 2222-2230, 2012.

[26] D. Nie, L. Wang, Y. Gao and D. Shen, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," In 13th international symposium on biomedical imaging (ISBI), IEEE, pp. 1342-1345, 2016

[27] P. Flandrin, "Time–Frequency Filtering Based on Spectrogram Zeros," Signal Processing Letters, IEEE, vol. 22, 2015.