

Limited receptive area neural classifier for recognition of swallowing sounds using continuous wavelet transform

Oleksandr Makeyev, *Associate Member, IEEE*, Edward Sazonov, *Member, IEEE*, Stephanie Schuckers, *Member, IEEE*, Paulo Lopez-Meyer, Ed Melanson, and Michael Neuman, *Member, IEEE*

Abstract— In this paper we propose a sound recognition technique based on the limited receptive area (LIRA) neural classifier and continuous wavelet transform (CWT). LIRA neural classifier was developed as a multipurpose image recognition system. Previous tests of LIRA demonstrated good results in different image recognition tasks including: handwritten digit recognition, face recognition, metal surface texture recognition, and micro work piece shape recognition. We propose a sound recognition technique where scalograms of sound instances serve as inputs of the LIRA neural classifier. The methodology was tested in recognition of swallowing sounds. Swallowing sound recognition may be employed in systems for automated swallowing assessment and diagnosis of swallowing disorders. The experimental results suggest high efficiency and reliability of the proposed approach.

I. INTRODUCTION

Many signal processing methods have been developed for analysis of non-stationary signals such as, for example, continuous wavelet transform (CWT). Conversion of a time domain signal into time-frequency domain increases the size of feature space from a one-dimensional signal to a two-dimensional “image”. Pattern recognition usually relies heavily on a few features extracted from the feature space and feature extraction algorithms represent a distinct area of research. We propose to apply LIRA-based image recognition technique to the “images” (scalograms) of the time-frequency decomposition of a sound instance [1]. This approach eliminates the need for a separate feature extraction algorithm.

Limited receptive area (LIRA) neural classifier was developed as a multipurpose image recognition system [2], [3] and tested with promising results in different image recognition tasks including: handwritten digit image recognition [4], micro device assembly [5], mechanically treated metal surface texture recognition [6], face recognition

[7], and micro work piece shape recognition [2].

In [1] we proposed a sound recognition technique combining the LIRA neural classifier and short-time Fourier transform (STFT). Sound recognition methodology based on combination of LIRA and STFT was validated with good results on recognition of swallowing sounds. Swallowing sound recognition is an important task in bioengineering that could be employed in systems for automated swallowing assessment and diagnosis of abnormally high rate of swallowing (aerophagia) [8], which is the primary mode of ingesting excessive amounts of air, and swallowing dysfunction (dysphagia) [9]-[12], that may lead to aspiration, choking, and even death, and represents a major problem in rehabilitation of stroke and head injury patients.

In current clinical practice videofluoroscopic swallow study (VFSS) is the gold standard for diagnosis of swallowing disorders. However, VFSS is not portable, time-consuming, and results in some radiation exposure. Therefore, various non-invasive methods are proposed for swallowing assessment based on evaluation of swallowing sounds, recorded by microphones and/or accelerometers and analyzed by digital signal processing techniques [9]-[12]. Swallowing sounds are caused by a bolus passing through pharynx. It is possible to use swallowing sounds to determine pharyngeal phase of the swallow and characteristics of the bolus [9].

Several techniques were proposed for automated detection of swallowing and breath sounds. In [10] an algorithm based on multilayer feed forward neural network was used for decomposition of tracheal sounds into swallowing and respiratory sound segments. The algorithm was able to detect 91.7% of swallows correctly for healthy subjects. In [11] a wavelet transform based filter with iterative sequences of multiresolution decomposition and reconstruction was used to differentiate swallowing sounds from breath sounds in healthy and dysphagic subjects. Ninety three percent of the swallowing sounds were detected correctly.

In a practical situation sound artifacts such as talking, throat clearing, and head movement that may be confused with swallowing and breath sounds decrease the efficiency of the recognition [12]. In [12] two sets of hybrid fuzzy logic committee neural networks (FCN) were proposed for recognition of dysphagic swallows, normal swallows and artifacts (speech, head movement). Swallows were detected by an ultra miniature accelerometer attached to the skin in

This work was supported in part by National Institutes of Health grant 5R21HL083052-02.

O. Makeyev, E. Sazonov, S. Schuckers, and P. Lopez-Meyer are with the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY 13699 USA (e-mail: mckehev@cias.clarkson.edu, esazonov@cias.clarkson.edu, sshuckers@cias.clarkson.edu, lopezmp@clarkson.edu).

E. Melanson is with the Center for Human Nutrition, University of Colorado Health Sciences Center, Denver, CO 80262 USA (e-mail: ed.melanson@uchsc.edu).

M. Neuman is with the Department of Biomedical Engineering, Michigan Technological University, Houghton, MI 49931 USA (e-mail: mneuman@mtu.edu).

the midline of the throat at the level of thyroid cartilage. Evaluation results revealed that FCN correctly identified 31 out of 33 dysphagic swallows, 24 out of 24 normal swallows, and 44 out of 45 artifacts. The ability to recognize swallow signal and eliminate artifacts with high accuracy is very important for development of home/tele-therapy biofeedback systems [13].

In this paper we demonstrate recognition of swallowing sounds using CWT in combination with the LIRA neural classifier and compare results with a similar approach using STFT [1].

II. METHODOLOGY

A. Data collection

Database of sounds was adopted from previous study [1].

Twenty sound instances were recorded for each of three classes of sounds (swallow, talking, head movement) with a commercially available miniature throat microphone (IASUS NT, IASUS Concepts Ltd.) located over laryngopharynx on a healthy subject without any history of swallowing disorder, eating or nutrition problems, or lower respiratory tract infection. An approval for this study was obtained from Institutional Review Board and the subject was asked to sign an informed consent form. To record the swallowing sound the subject was asked to consume water in boluses of arbitrary size.

Utilized throat microphone converts vibration signals from the surface of the skin rather than pick up waves of sound pressure, thus reducing the ambient noise. The microphone also picks up such artifacts as head movements and talking that should not be confused with swallowing sounds. For head movement artifact recording the subject was asked to turn his head to a side and back. To record speech the subject was asked to say the word “Hello”. Sound signals for each class were amplified and recorded with a sampling rate of 44100 Hz.

A fourth class of outlier sounds that consisted of random segments of music recordings was introduced to demonstrate the ability of the neural classifier to reject sounds with weak intra-class similarity and no similarity with other three classes.

B. Data preprocessing

Swallowing, head movement, and talking sounds were extracted from the recordings in segments of 65536 samples (approximately 1.5 s) each using the following empiric algorithm: beginning and end of each sound were found using a threshold set above the background noise level, center of mass was calculated for each sound and used to center the corresponding sound instance in the recognition window.

Morlet mother wavelet with wavenumber of 6, 7 octaves and 16 suboctaves [14] was used to obtain scalograms of sound instances. To compare pattern recognition accuracy on

time-frequency decompositions produced by CWT and STFT [1] the following processing was applied to the scalograms: a mirror image of the scalograms across abscissa was created and combined with the original; the resulting image was resized to 256x256 pixels using bicubic interpolation. Fig. 1. shows examples of scalogram images. Eighty grayscale scalogram images (20 for each of 4 classes) compose the image database that was used in training and validation.

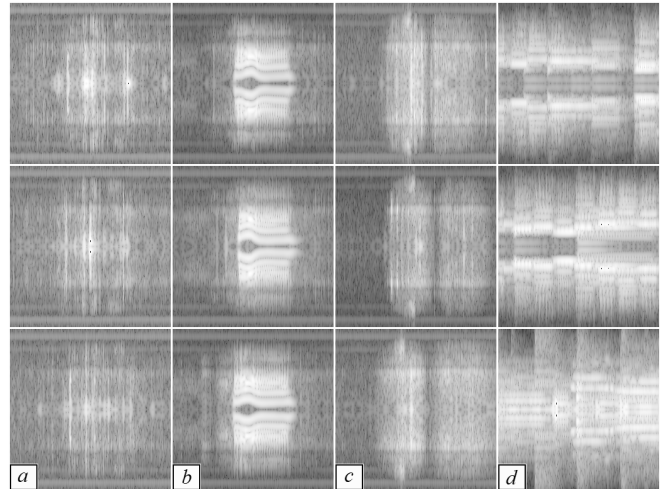


Fig. 1. Examples of scalograms of (columns): a) swallowing sounds, b) talking, c) head movements, d) outlier sounds.

C. LIRA neural classifier

The LIRA neural classifier is a four-layer perceptron [15] that consists of *S*-layer, *I*-layer, *A*-layer and *R*-layer (Fig. 2).

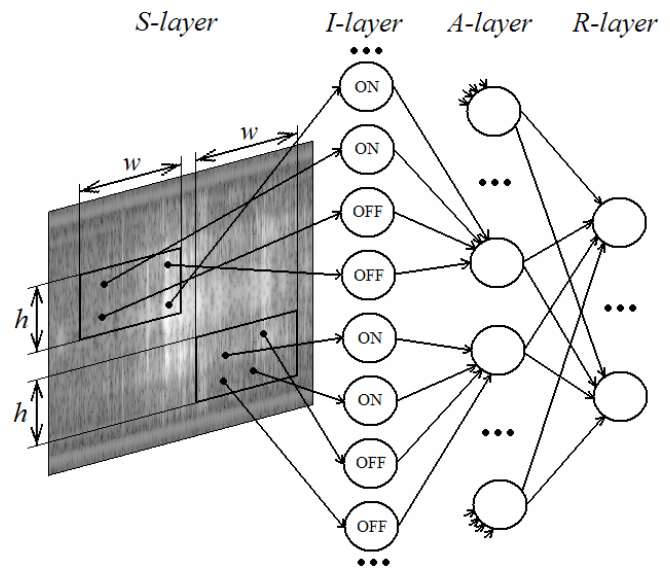


Fig. 2. Structure of the LIRA neural classifier.

Sensor *S*-layer corresponds to input image. Associative *A*-layer is connected to the *S*-layer through the intermediate *I*-layer with randomly selected non-trainable connections. The

set of these connections can be considered as a feature extractor. Intermediate I -layer that consists of ON- and OFF-neurons is designed to work with grayscale images. The input of each I -layer neuron is connected to one neuron of the S -layer and the output is connected to the input of one neuron of the A -layer. All the I -layer neurons connected to one A -layer neuron form the group of this A -layer neuron. For example, in Fig. 2 the group of four I -layer neurons, two ON-neurons and two OFF-neurons, corresponds to one A -layer neuron. Reaction R -layer contains neurons that correspond to output classes of the LIRA classifier. Each neuron of the A -layer is connected to all the neurons of the R -layer. The weights of these connections are modified during the classifier training.

The fixed set of connections between the S -layer and the A -layer is created with the following procedure repeated for all the A -layer neurons: the window of height h and width w is randomly located in the S -layer; inputs of a group of I -layer ON- and OFF-neurons are connected to random neurons within the window $h \cdot w$ of the S -layer and outputs are connected to the A -layer neuron; the thresholds of ON- and OFF-neurons are selected randomly from the range $[0, b_{max}]$, where b_{max} is the maximal brightness of image pixels.

Each input image defines unique activations of the A -layer neurons. After the set of connections between the S -layer and the A -layer is created the binary vector that represents outputs of associative neurons is calculated for each image of training and validation sets in accordance to following rules:

1. The output of the ON-neuron is equal to 1 if its input value is larger than its threshold and it is equal to 0 in the opposite case. The output of the OFF-neuron is equal to 1 if its input value is smaller than its threshold and it is equal to 0 in the opposite case.

2. The output of the A -layer neuron is equal to 1 if outputs of all the neurons of its I -layer group are equal to 1 and it is equal to 0 in the opposite case.

These binary vectors will be used during training and recognition procedures.

Training and recognition procedures of the LIRA neural classifier are similar to the ones of the perceptron [15]. The training process is carried out iteratively. In each training cycle all the images of the training set are presented to the neural classifier. During the recognition process all the images of the validation set are presented to the neural classifier.

Performance of a LIRA neural classifier can be improved with implementation of distortions of input images during training and recognition [2]. In our experiments we used different combinations of horizontal, vertical and bias image translations.

A detailed description of the LIRA neural classifier can be found in [2], [3].

III. RESULTS

In our experiments we used holdout cross-validation, i.e.

the validation set for each class was chosen randomly from the database and the rest of the database was used for training. In each experiment we performed 50 runs of the holdout cross-validation to obtain statistically reliable results. A new set of connections between the S -layer and the A -layer and a new division into the training and validation sets were created for each run. The number of images in training set varied from two to ten.

Mean recognition rate was calculated from the mean number of errors for one run and the total number of images in the validation set. Comparison of recognition rates obtained with combination of LIRA with CWT and STFT for various numbers of images in training and validation sets is presented in Table 1.

Table 1. Comparison of recognition rates for combination of LIRA with CWT and STFT

T/V *	Mean recognition rate (%)		P -value for paired t -test for mean recognition rate	95% lower bound for mean difference
	CWT	STFT		
2/18	99.31	98.75	0.015	0.14
4/16	99.84	99.63	0.035	0.021
6/14	99.93	99.71	0.068	-0.022
8/12	100	99.92	0.08	-0.014
10/10	100	100	-	-

* T is the size of the training set for each class, V is the size of the validation set for each class.

The following set of LIRA parameters was used during all the experiments: window $h \cdot w$ width $w = 10$, height $h = 10$; the number of training cycles is 30; the number of ON-neurons in the I -layer neuron group that corresponds to one A -layer neuron is 3, the number of OFF-neurons is 5; 8 distortions for training including ± 1 pixel horizontal, vertical and bias image translations and 4 distortions for recognition including ± 1 pixel horizontal and vertical image translations; the total number of associative neurons $N = 512,000$.

Paired t -test [16] for mean recognition rate was used to evaluate significance of difference in recognition rates for CWT and STFT with null hypothesis of no difference in recognition rates and alternative of mean recognition rate for CWT being higher than the one for STFT. P -values and 95% lower bounds for mean difference are presented in Table 1.

IV. DISCUSSION

The main advantage of CWT over STFT is the tiling of the resolution. Time-frequency resolution of STFT is constant which results in the uniform tiling of time-frequency plane with a rectangular cell of fixed dimensions. For CWT the time-frequency resolution varies according to the frequency of interest. CWT resolution is finer at higher frequencies at the cost of a larger frequency window while the area of each cell is constant. Hence, CWT can discern individual high frequency features located close to each other in the signal, whereas STFT smears such high

frequency features occurring within its fixed width time window [14]. This advantage of CWT is reflected in the experimental results. Results of paired *t*-test indicate that we can reject the null hypothesis for alternative of mean recognition rate for CWT being higher than the one for STFT at the level of significance ranging from 92% for eight images in training set to 98.5% for two images in the training set, indicating a statistically significant improvement in the recognition rate.

In our experiments we set the parameter values to maximize efficiency of the LIRA neural classifier. The amount of time needed for one run of classifier coding, training and recognition with the set of parameters presented in Section IV is approximately 2 min (90 s for coding, 30 s for training and 1 s for recognition) on a computer equipped with AMD Athlon 64 X2 4400+ Dual Core processor and 2.00 GB of RAM. This allows the recognition of swallowing instances in sound streams to be performed with the highest recognition rate on a personal computer in a real-time.

Obtained results suggest high efficiency and reliability of the proposed method, though tests on a larger database would be needed for a conclusive proof. The demonstrated results are based on an individual recognition model derived for a single subject. Future work includes further experiments on individual and general (group) models. An important advantage of the proposed method is utilization of a double-redundant approach to identification of significant features. First, time-frequency decomposition method provides a redundant description of a sound instance, therefore increasing chances for random selection of a significant feature. Second, randomly assigned redundant connections between the sensor and associate layers ensure multiplicity of extracted features and good description of an image without prior knowledge about the image content. Features that do not provide useful information for separation of classes will not obtain significant weights during training. The proposed methodology presents a novel deviation from the traditional approach of using small sets of empirically-selected statistics as features in sound recognition.

CONCLUSION

In this paper we propose a sound recognition technique based on the limited receptive area (LIRA) neural classifier and continuous wavelet transform (CWT). The proposed technique works by applying a LIRA-based image recognition system to the scalograms of sound instances.

The suggested methodology is tested in recognition of four classes of sounds that correspond to swallowing sounds, talking, head movements and outlier sounds. Experimental results suggest high efficiency and reliability of the proposed method as well as its superiority over the previously proposed method combining LIRA with short-time Fourier transform (STFT).

The proposed method may be employed in systems for

automated swallowing assessment and diagnosis of swallowing disorders and has potential for application to other sound recognition tasks.

ACKNOWLEDGMENT

The authors gratefully acknowledge E. Kussul and T. Baidyk, UNAM, Mexico, for the constructive discussions and helpful comments.

REFERENCES

- [1] O. Makeyev, E. Sazonov, S. Schuckers, E. Melanson, M. Neuman, "Limited receptive area neural classifier for recognition of swallowing sounds using short-time Fourier transform", in *Proc. International Joint Conference on Neural Networks IJCNN'2007*, Orlando, USA, 2007, pp. 6, accepted for publication.
- [2] E. Kussul, T. Baidyk, D. Wunsch, O. Makeyev, A. Martín, "Permutation coding technique for image recognition systems," *IEEE Trans Neural Networks*, vol. 17/6, pp. 1566-1579, 2006.
- [3] E. Kussul, T. Baidyk, D. Wunsch, O. Makeyev, A. Martín, "Image recognition systems based on random local descriptors," in *Proc. International Joint Conference on Neural Networks IJCNN'2006*, Vancouver, Canada, 2006, pp. 4722-4727.
- [4] E. Kussul, T. Baidyk, "Improved method of handwritten digit recognition tested on MNIST database," *Image and Vision Computing*, vol. 22, pp. 971-981, 2004.
- [5] T. Baidyk, E. Kussul, O. Makeyev, A. Caballero, L. Ruiz, G. Carrera, G. Velasco, "Flat image recognition in the process of microdevice assembly," *Pattern Recogn Lett.*, vol. 25/1, pp. 107-118, 2004.
- [6] O. Makeyev, T. Baidyk, A. Martín, "Limited receptive area neural classifier for texture recognition of metal surfaces," in *Proc. IFIP WCC AI2006*, Santiago de Chile, Chile, 2006, pp. 10.
- [7] E. Kussul, T. Baidyk, M. Kussul, "Neural network system for face recognition," in *Proc. IEEE International Symposium on Circuits and Systems, ISCAS*, Vancouver, Canada, 2004, vol.V, pp. V-768-V-771.
- [8] A. K. Limdi, M. J. McCutcheon, E. Taub, W. E. Whitehead, E. W. Cook, "Design of a microcontroller-based device for deglutition detection and biofeedback," in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Seattle, USA, 1989, vol. 5, pp. 1393-1394.
- [9] T. Nakamura, Y. Yamamoto, H. Tsugawa, "Measurement system for swallowing based on impedance pharyngography and swallowing sound," in *Proc. 17th IEEE Instrumentation and Measurement Technology Conference*, Baltimore, Maryland, USA, 2000, vol. 1, pp. 191-194.
- [10] M. Aboofazeli, Z. Moussavi, "Automated classification of swallowing and breath sounds," in *Proc. 26th Annual International Conference of the Engineering in Medicine and Biology Society*, San Francisco, California, USA, 2004, vol. 2, pp. 3816-3819.
- [11] M. Aboofazeli, Z. Moussavi, "Automated Extraction of Swallowing Sounds Using a Wavelet-Based Filter," in *Proc. 28th Annual International Conference of the Engineering in Medicine and Biology Society*, New York, New York, USA, 2006, pp. 5607-5610.
- [12] A. Das, N. P. Reddy, J. Narayanan, "Hybrid fuzzy-neural committee networks for recognition of swallow acceleration signals", *Computer Methods and Programs in Biomedicine*, vol. 64, pp. 87-99, 2000.
- [13] N. P. Reddy, V. Gupta, A. Das, R. N. Unnikrishnan, G. Song, D. L. Simcox, H. P. Reddy, S. K. Sukthankar, E. P. Canilang, "Computerized biofeedback system for treating dysphagic patients for traditional and teletherapy applications," in *Proc. International Conference on Information Technology Application in Biomedicine ITAB'98*, Piscataway, New Jersey, USA, 1998, 100-104.
- [14] P. S. Addison, "The illustrated wavelet transform handbook". Institute of Physics Publishing, Bristol, 2002.
- [15] F. Rosenblatt, "Principles of neurodynamics". Spartan books, New York, 1962.
- [16] D. C. Montgomery, "Design and analysis of experiments", Wiley, Hoboken, 2004.