Recognition of Swallowing Sounds Using Time-Frequency Decomposition and Limited Receptive Area Neural Classifier

O. Makeyev[1], E. Sazonov[2], S. Schuckers[3], P. Lopez-Meyer[4], T. Baidyk[5], E. Melanson[6] and M. Neuman[7]

Abstract In this paper we propose a novel swallowing sound recognition technique based on the limited receptive area (LIRA) neural classifier and time-frequency decomposition. Time-frequency decomposition methods commonly used in sound recognition increase dimensionality of the signal and require steps of feature selection and extraction. Quite often feature selection is based on a set of empirically chosen statistics, making the pattern recognition dependent on the intuition and skills of the investigator. A limited set of extracted features is then presented to a classifier. The proposed method avoids the steps of feature selection and extraction by delegating them to a limited receptive area neural (LIRA) classifier. LIRA neural classifier utilizes the increase in dimensionality of the signal to create a large number of random features in the time-frequency domain that assure a good description of the signal without prior assumptions of the signal properties. Features that do not provide useful information for separation of classes do not obtain significant weights during classifier training. The proposed methodology was tested on the task of recognition of swallowing sounds with two different algorithms of time-frequency decomposition, short-time Fourier

---

1 Department of Electrical and Computer Engineering, Clarkson University, Potsdam NY USA, mckehev@cias.clarkson.edu

2 Department of Electrical and Computer Engineering, Clarkson University, Potsdam NY USA, esazonov@cias.clarkson.edu

3 Department of Electrical and Computer Engineering, Clarkson University, Potsdam NY USA, sschuckers@cias.clarkson.edu

4 Department of Electrical and Computer Engineering, Clarkson University, Potsdam NY USA, lopezmp@cias.clarkson.edu

5 Center of Applied Research and Technological Development, National Autonomous University of Mexico, Mexico City Mexico, tbaidyk@servidor.unam.mx

6 Center for Human Nutrition, University of Colorado Health Sciences Center, Denver CO USA, ed.melanson@uchsc.edu

7 Department of Biomedical Engineering, Michigan Technological University, Houghton MI USA, mneuman@mtu.edu

transform (STFT) and continuous wavelet transform (CWT). The experimental results suggest high efficiency and reliability of the proposed approach.

1 Introduction

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) [1], which is the primary mode of ingesting excessive amounts of air, and swallowing dysfunction (dysphagia) [2]-[5], that may lead to aspiration, choking, and even death. Dysphagia 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 a time-consuming procedure performed only in a clinical setting. VFSS also results in some radiation exposure. Therefore, various non-invasive methods are proposed for swallowing assessment based on evaluation of swallowing signals, recorded by microphones and/or accelerometers and analyzed by digital signal processing techniques [2]-[5]. 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 [2].

Signal processing for swallowing sound detection is usually based on a series of steps such as decomposition of the raw signal, selection and extraction of features followed by pattern recognition. Many time-frequency decomposition methods have been developed for analysis of non-stationary signals including sounds such as, for example, short-time Fourier transform (STFT) and 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 power spectrum such as a spectrogram in case of STFT or a scalogram in case of CWT. Typically, the feature space generated by time-frequency decomposition is reduced to a few significant coefficients by computing various empirically chosen statistics for time and frequency windows, thus making pattern recognition dependent on a small set of extracted features. Selection of statistics and window parameters is left to the investigator that has to make sure that selected features match well with the signal characteristics. For example, in [3] an algorithm based on multilayer feed forward neural network was used for decomposition of tracheal sounds into swallowing and respiratory sound segments. Three features (root mean square, the average power of the signal over 150-450 Hz, and waveform fractal dimension) of the signal were used as inputs of the neural network. The algorithm was able to detect 91.7% of swallows correctly for healthy subjects. In a practical situations artifacts such as talking, throat clearing, and head movement may be confused with swallowing and breath decreasing the efficiency of the recognition [4], [5]. In [4] acceleration at the throat, during swallowing and coughing, was measured by an ultra miniature accelerometer placed on the skin at the level of thyroid cartilage. Two sets of neural networks were used to recognize and classify acceleration patterns due to swallowing and coughing in normal and dysphagic subjects. A set of features that include peak to peak amplitudes, slopes, mean frequency, number of zero crossings, and mean power were used as inputs of two multilayer feed forward neural networks. The recognition rate for acceleration patterns due to swallowing and coughing in normal and dysphagic subjects was 100% and the recognition rate for acceleration patterns due to swallowing in normal, mildly dysphagic, moderately dysphagic, and severely dysphagic subjects was 93%. In [5] 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 at the level of thyroid cartilage. Five features (number of zero crossings, average power, average frequency, maximum power, and frequency at maximum power) were extracted from the filtered signal, normalized, and used as inputs of the FCN. 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 [6].

In this paper we propose a novel sound recognition technique based on time-frequency decomposition and limited receptive area (LIRA) neural classifier which incorporates the feature selection and extraction steps. LIRA neural classifier was developed as a multipurpose image recognition system [7] and tested with promising results in different image recognition tasks including: handwritten digit image recognition [8], micro device assembly [9], mechanically treated metal surface texture recognition [10], face recognition [11], and micro work piece shape recognition [7]. A distinctive feature of the LIRA neural classifier is utilization of a large number of random features. Features that do not provide useful information for separation of classes do not obtain significant weights during training. We propose to apply LIRA-based image recognition technique to the “images” of time-frequency decomposition spectrums obtained by STFT [12] and CWT [13]. Utilization of a large number of random features in the time-frequency domain assures a good description of the signal without prior assumptions of the signal properties and eliminates the need for a separate feature selection and extraction algorithms. We demonstrate recognition of swallowing sounds using the proposed approach with two different time-frequency decomposition algorithms, i.e. using LIRA neural classifier in combination with STFT and CWT, and compare the obtained results. The paper is organized as follows: In section 2 we present the methodology including detailed description of the data collection process, data preprocessing, and LIRA neural classifier. In section 3 we present the experimental results. Discussion and conclusions are presented in sections 4 and 5 correspondingly.

2 Methodology

2.1 Data Collection

Commercially available miniature throat microphone (IASUS NT, IASUS Concepts Ltd.) located over laryngopharynx was used during the data collection process. Throat microphones convert vibration signals from the surface of the skin rather than pick up waves of sound pressure, thus reducing the ambient noise. Throat microphones also pick up such artifacts as head movements and talking that should not be confused with swallowing sounds.

Twenty sound instances were recorded for each of three classes of sounds (swallow, talking, head movement) for 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. For head movement artifact recording the subject was asked to turn his head to a side and back. To record speech artifact 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.

2.2 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. The same segmentation was used for both algorithms of time-frequency decomposition.

2.2.1 STFT

Spectrograms of each segment were calculated with a window of 512 samples extracted using a Hanning window algorithm and processed by STFT with 50% window overlap. Due to limited signal bandwidth higher frequencies do not contain significant energy of the original time domain signal and can be eliminated from the spectrogram. Truncating the spectrogram from 512x256 pixels to 256x256 pixels preserves most of the signal energy and eliminates insignificant harmonics. Figure 1 shows examples of spectrogram images. Eighty grayscale spectrogram images (20 for each of 4 classes) compose the first image database that was used in training and validation.

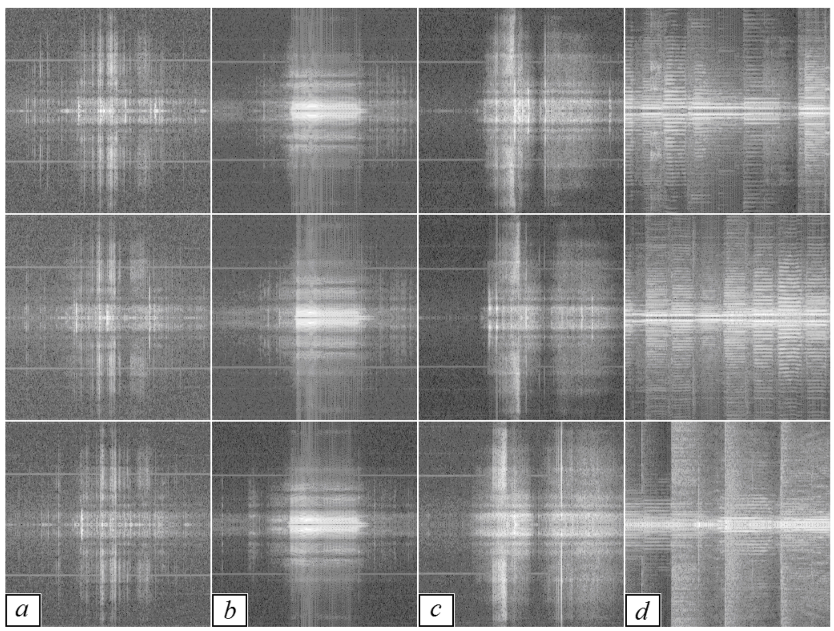

Figure 1 Examples of spectrograms of (columns): a) swallowing sounds b) talking, c) head movements, d) outlier sounds.

2.2.2 CWT

Morlet mother wavelet with wavenumber of 6, 7 octaves and 16 sub-octaves [14] was used to obtain scalograms of sound instances. To compare pattern recognition accuracy on time-frequency decompositions produced by CWT and STFT 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. Figure 2 shows examples of scalogram images. The same set of sound instances was used to create figure 1 and figure 2 allowing the direct visual comparison to be drawn. Eighty grayscale scalogram images (20 for each of 4 classes) composed the second image database that was used in training and validation.

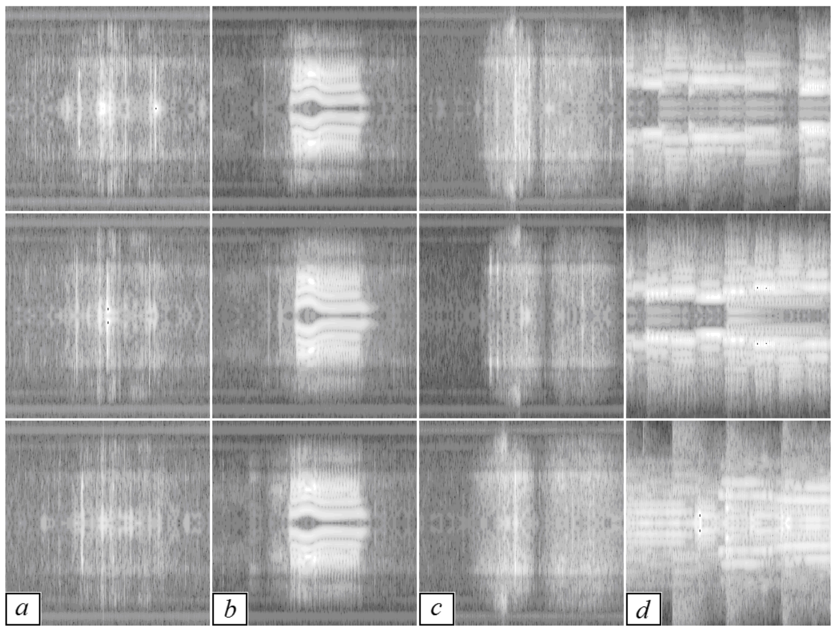

Figure 2 Examples of scalograms of (columns): a) swallowing sounds b) talking, c) head movements, d) outlier sounds.

2.3 LIRA Neural Classifier

LIRA neural classifier was developed on the basis of the Rosenblatt perceptron [15]. The three-layer Rosenblatt perceptron consists of the sensor *S*-layer, associative *A*-layer and the reaction *R*-layer. The first *S*-layer corresponds to the input image or in our case, a spectrogram or a scalogram. The second *A*-layer corresponds to the feature extraction subsystem. The third *R*-layer represents the system's output. Each neuron of this layer corresponds to one of the output classes.

The associative layer A is connected to the sensor layer S with randomly selected, non-trainable connections. The set of these connections can be considered as a feature extractor.

The A-layer consists of 2-state neurons; their outputs can be equal either to 1 (active state) or to 0 (non-active state). 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.

To perform pattern recognition on grayscale scalograms and spectrograms an additional 2-state neuron layer was introduced between the S-layer and the A-layer. We term it the I-layer (intermediate layer). The structure of the LIRA neural classifier is presented in figure 3.

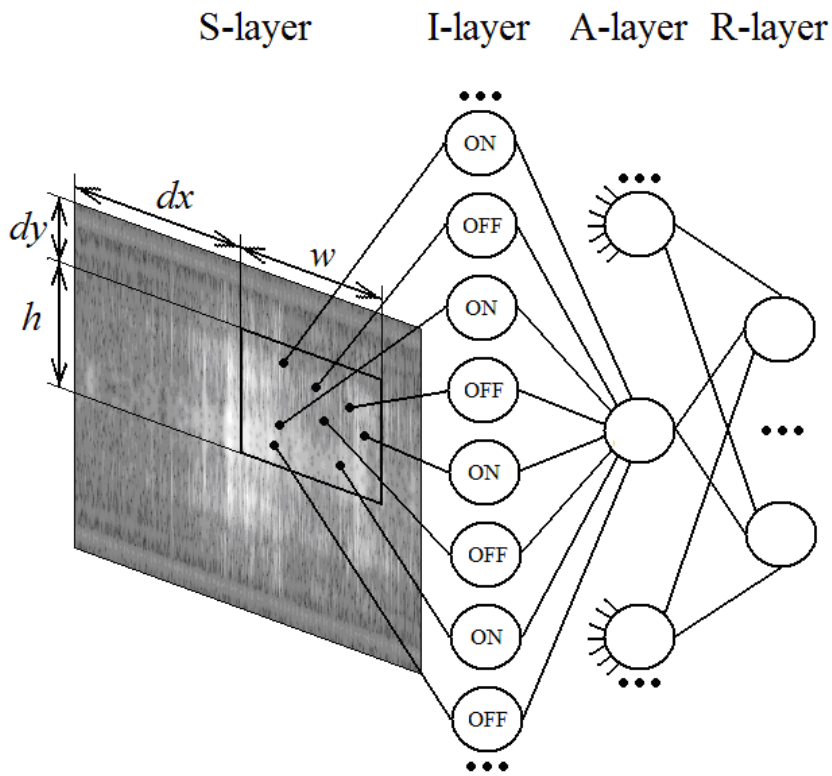

Figure 3 Structure of the LIRA neural classifier.

2.3.1 Coding Procedure

Each input time-frequency decomposition defines unique activations of the *A*-layer neurons. The binary vector that corresponds to the associative neuron activations is termed the binary code $A = (a_1, \ldots, a_N)$, where N is the number of the *A*-layer neurons. The procedure that transforms input time-frequency data into corresponding binary vector A is termed the image coding.

We connect each *A*-layer neuron to *S*-layer neurons randomly selected from a randomly generated window of height h and width w that is located in the *S*-layer (figure 3).

The distances dx and dy are random numbers selected from the ranges: dx from $[0, W_S - w]$ and dy from $[0, H_S - h]$, where W_S and H_S stand for width and height of the *S*-layer. We create the associative neuron masks that represent the positions of connections of each *A*-layer neuron with neurons of the window $h \cdot w$. The procedure of random selection of connections is used to design the mask of *A*-layer neurons. This procedure starts with the selection of the upper left corner of the window $h \cdot w$ in which all connections of the associative neuron are located.

The following formulas are used:

$dx_i = random_i\,(W_S - w)$,

$dy_i = random_i\,(H_S - h)$,

where i is the position of a neuron in associative layer A, $random_i\,(z)$ is a random number that is uniformly distributed in the range $[0, z]$. After that position of each connection within the window $h \cdot w$ is defined by the pair of numbers:

$x_{ij} = random_{ij}\,(w)$,

$y_{ij} = random_{ij}\,(h)$,

where j is the number of the connection with the *S*-layer.

Absolute coordinates of a connection to the *S*-layer are defined as:

$X_{ij} = x_{ij} + dx_i$,

$Y_{ij} = y_{ij} + dy_i$.

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. There are two types of *I*-layer neurons: ON-neurons and OFF-neurons. The outputs of ON- and OFF- neurons are computed according to the formula:

$$ON_i = \begin{cases} 1, b_i > \theta_i \\ 0, b_i \le \theta_i \end{cases},$$

$$OFF_j = \begin{cases} 1, b_j < \theta_j \\ 0, b_j \ge \theta_j \end{cases},$$

where ON_i and OFF_j are the outputs of ON-neuron i and OFF-neuron j correspondingly, θ_i and θ_j are the thresholds of ON-neuron i and OFF-neuron j correspondingly, and b_i and b_j are the values of brightness of the image pixels that correspond to ON-neuron i and OFF-neuron j correspondingly. Thresholds θ_i and θ_j are selected randomly from the range $[0, b_{max}]$, where b_{max} is maximal brightness of the image pixels. For example, in figure 3 the group of eight *I*-layer neurons, four ON-neurons and four OFF-neurons, corresponds to one *A*-layer neuron. The *i*-th neuron of the *A*-layer is active ($a_i = 1$) only if outputs of all the neurons of its *I*-layer group are equal to 1 and is non-active ($a_i = 0$) in the opposite case.

After execution of the coding procedure each signal has an associated binary code that is to be used during the training and recognition procedures.

2.3.2 Training procedure

The weights of all the connections between neurons of the *A*-layer and the *R*-layer are initialized to zero prior to execution of the training procedure. Next, training is performed using the following algorithm:

Step 1. Calculation of excitation.

A binary code *A* is presented to the LIRA neural classifier. *R*-layer neuron excitations E_i are computed according to the formula:

$$E_i = \sum_{j=1}^{N} a_j \cdot w_{ji},$$

where E_i is the excitation of the *i*-th neuron of the *R*-layer, a_j is the output signal (0 or 1) of the *j*-th neuron of the *A*-layer, w_{ji} is the weight of the connection between the *j*-th neuron of the *A*-layer and the *i*-th neuron of the *R*-layer.

Step 2. Excitation adjustment.

Excitation adjustment is performed after calculation of the neuron excitations of the *R*-layer. The excitation E_c of the *R*-layer neuron that corresponds to the correct class *c* is recalculated according to the formula:

$$E_c^* = E_c \cdot (1 - T_E),$$

where $0 \le T_E \le 1$ determines the reserve of excitation the neuron that corresponds to the correct class must have. In our experiments the value T_E varied from 0.1 to 0.5.

Next, the neuron with the largest excitation is selected. This winner neuron represents the recognized class *j*.

Step 3. Adjustment of weights.

If the winning neuron corresponds to the correct class *c* ($j = c$) then no modification of weights is needed. If $j \ne c$ then following modification of weights is performed:

$$w_{ic}(t+1) = w_{ic}(t) + a_i,$$

$$w_{ij}(t+1) = w_{ij}(t) - a_i,$$

where $w_{ij}(t)$ and $w_{ij}(t+1)$ are the weights of the connection between the *i*-th neuron of the *A*-layer and the *j*-th neuron of the *R*-layer before and after modification, a_i is the output signal (0 or 1) of the *i*-th neuron of the *A*-layer.

The training process is carried out iteratively. In each training cycle all the binary vectors of the training set are presented to the neural classifier.

Image recognition performance of a LIRA neural classifier can be improved with implementation of distortions of input images during training and recognition [7]. In our experiments we used different combinations of horizontal, vertical and bias translations of the spectrograms and scalograms.

2.3.3 Recognition Procedure

Distortions of the time-frequncy data have been used both for training and recognition. There is an essential difference between implementation of distortions for training and recognition. In the training process each distortion is considered as an independent new data point in the training set. Distortions introduced during the recognition process are not treated as independent data. Instead a decision-making rule is applied in order to make a decision about the class label assignment based on the original data and all of its distortions. The decision-making rule consists in calculation of the R-layer neuron excitations for all the distortions sequentially:

$$E_i = \sum_{k=0}^{d} \sum_{j=1}^{N} a_{kj} \cdot w_{ji},$$

where E_i is the excitation of the i-th neuron of the R-layer, a_{kj} is the output signal (0 or 1) of the j-th neuron of the A-layer for the k-th distortion, w_{ji} is the weight of the connection between the j-th neuron of the A-layer and the i-th neuron of the R-layer, d is the number of applied distortions (case $k = 0$ corresponds to the initial data).

The neuron with the largest excitation (winner neuron) represents the recognized class.

3 Results

The proposed methodology was tested on two datasets composed correspondingly of spectrograms and scalograms of sound instances. Each database contained eighty 256x256 pixels grayscale images (20 for each of 4 classes).

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 sounds in training and validation sets for each class equals to ten, i.e. database is divided in half.

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

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

Number of associative neurons	Mean recognition rate (%)		*P*-value for paired *t*-test for mean recognition rate	95% lower bound for mean difference
	CWT	STFT		
1,000	85.3	81.75	0.02	0.72
2,000	96.5	94.25	0.002	1.039
4,000	99.6	98.1	< 0.001	0.926
8,000	100	99.85	0.042	0.0078

The following set of LIRA parameters was used during all the experiments: window $h \cdot w$ width w = 10, height h = 10; reserve of excitation T_E = 0.3; 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 translations and 4 distortions for recognition including ±1 pixel horizontal and vertical translations.

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.

4 Discussion

Obtained results suggest high efficiency and reliability of the proposed method, though tests on a larger database would be needed for a conclusive proof. 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 associative layers ensure multiplicity of extracted random features. The number of extracted random features which is equal to the total number of associative neurons is a crucial parameter of the system. This is reflected in the experimental results presented in Table 1. This number should be sufficiently large to create a detailed description of a sound instance providing a basis for further classification. At the same time large number of associative neurons results in computational burden that may pose additional problems for consumer applications, for example, in case of a practical implementation of our system in form of a wearable device extra large number of associative neurons may result in computational burden that would exceed the capabilities of a typical low-power embedded processor. In our case, the amount of time needed for one run of classifier coding, training and recognition with the set of parameters presented in Section IV and 8,000 of associative neurons is approximately 75 sec (72 s for coding, 2 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.

We attribute the higher accuracy achieved in classification of CWT data to 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 [15]. 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 for all numbers of associative neurons (P-value < 0.05).

The method presented here achieves similar or higher accuracy compared to previously published methods. The advantage of this approach is that our method of feature extraction is automated. This has two main advantages. The first is that our method is not necessarily tailored to a specific collected dataset. That is, often in applications, features that are chosen manually from one dataset may achieve high performance for that dataset, but are not generalizable to the underlying application. The second is that manually chosen features may not achieve the best performance because potentially useful features for classification may have been overlooked. More research is needed to assess the generalizability and performance with this method of classifying swallowing sounds compared with other approaches.

5 Conclusions

In this paper we propose a novel swallowing sound recognition technique based on the limited receptive area (LIRA) neural classifier and time-frequency decomposition. The proposed technique works by applying a LIRA-based multipurpose image recognition system to the time-frequency decomposition spectrums of sound instances with extraction of a large number of random features. Features that do not provide useful information for separation of classes do not obtain significant weights during training. This approach eliminates the need for empirical feature selection and therefore simplifies design of pattern recognition systems for non-stationary signals such as swallowing sounds.

The proposed methodology is tested with two different algorithms of time-frequency decomposition, short-time Fourier transform (STFT) and continuous wavelet transform (CWT), 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 superiority of combination of LIRA with CWT over the combination of LIRA with STFT.

The proposed multipurpose sound recognition technique may be employed in systems for automated swallowing assessment and diagnosis of swallowing disorders and has potential for application to other sound recognition tasks.

6 Acknowledgment

This work was supported in part by National Institutes of Health grant R21HL083052-02.
The authors gratefully acknowledge E. Kussul, National Autonomous University of Mexico (UNAM), for the constructive discussions and helpful comments.

References

1. Limdi, A.K., McCutcheon, M.J., Taub, E., Whitehead, W.E., Cook, E.W.: Design of a microcontroller-based device for deglutition detection and biofeedback. In: Proceedings of 11th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Seattle, USA pp. 1393-1394 (1989).
2. Nakamura, T., Yamamoto, Y., Tsugawa, H.: Measurement system for swallowing based on impedance pharyngography and swallowing sound. In: Proceedings of 17th IEEE Instrumentation and Measurement Technology Conference, Baltimore, Maryland, USA pp. 191-194 (2000).
3. Aboofazeli, M., Moussavi, Z.: Automated classification of swallowing and breath sounds. In: Proceedings of 26th Annual International Conference of the Engineering in Medicine and Biology Society, San Francisco, California, USA pp. 3816-3819 (2004).
4. Prabhu, D.N.F., Reddy, N.P., Canilang, E.P.: Neural networks for recognition of acceleration patterns during swallowing and coughing. In: Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Baltimore, Maryland, USA pp. 1105-1106 (1994).

5. Das, A., Reddy, N.P., Narayanan, J.: Hybrid fuzzy-neural committee networks for recognition of swallow acceleration signals. Comput. Meth. Prog. Bio. Vol. 64, pp. 87-99 (2000).
6. Reddy, N.P., Gupta, V., Das, A., Unnikrishnan, R.N., Song, G., Simcox, D.L., Reddy, H.P., Sukthankar, S.K., Canilang, E.P.: Computerized biofeedback system for treating dysphagic patients for traditional and teletherapy applications. In: Proceedings of International Conference on Information Technology Application in Biomedicine ITAB'98, Piscatway, New Jersey, USA pp. 100-104 (1998).
7. Kussul, E., Baidyk, T., Wunsch, D., Makeyev, O., Martín, A.: Permutation coding technique for image recognition systems. IEEE Trans. Neural Networks Vol. 17 pp. 1566-1579 (2006).
8. Kussul, E., Baydik, T.: Improved method of handwritten digit recognition tested on MNIST database. Image Vision Comput. Vol. 22 pp. 971-981 (2004).
9. Baidyk, T., Kussul, E., Makeyev, O., Caballero, A., Ruiz, L., Carrera, G., Velasco, G.: Flat image recognition in the process of microdevice assembly. Pattern Recogn. Lett. Vol. 25 pp. 107-118 (2004).
10. Makeyev, O., Sazonov, E., Baidyk, T., Martín, A.: Limited receptive area neural classifier for texture recognition of mechanically treated metal surfaces, Neurocomputing, Vol. 71 pp. 1413-1421 (2008).
11. Kussul, E., Baidyk, T., Kussul, M.: Neural network system for face recognition. In: Proceedings of IEEE Internatrional Symposium on Circuits and Systems, ISCAS'2004, Vancouver, Canada pp. 768-771 (2004).
12. Makeyev, O., Sazonov, E., Schuckers, S., Melanson, E., Neuman, M.: Limited receptive area neural classifier for recognition of swallowing sounds using short-time Fourier transform. In: Proceedings of International Joint Conference on Neural Networks IJCNN'2007, Orlando, USA pp. 1417.1-1417.6 (2007).
13. Makeyev, O., Sazonov, E., Schuckers, S., Lopez-Meyer, P., Melanson, E., Neuman, M. Limited receptive area neural classifier for recognition of swallowing sounds using continuous wavelet transform. In: Proceedings of 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC'2007, Lyon, France pp. 3128-3131 (2007).
14. Addison, P.S.: The illustrated wavelet transform handbook. Institute of Physics Publishing, Bristol (2002).
15. Rosenblatt, F.: Principles of neurodynamics. Spartan books, New York (1962).
16. Montgomery, D.C.: Design and analysis of experiments. Wiley, Hoboken (2004).