J. King Saud Univ., Vol. 20, *Eng. Sci.* (2), pp. 113-121, Riyadh (2008/1429H.)

Voice-activated M2M Communication over GSM Network

Abdulmohsen Alheraish, Fahad Alshehry, Saleh Alshebeili and Mohammed Abu ElEla
Department of Electrical Engineering, College of Engineering,
King Saud University, P.O. Box 800, Riyadh 11421,
Saudi Arabia

(Received on 22 November 2006; accepted for publication on 28 April 2007)

Keywords: Voice-based control device, Machine-to-machine, Mobile, Neural network.

Abstract. Wireless M2M (machine-to-machine) has emerged recently as a new technology that allows machines to transmit or receive data remotely over mobile communication networks such as GSM. The conventional wireless M2M systems available on the market today allow inputting data from mobile phone keypad. In this paper, a novel prototype system design is developed that would use voice to establish M2M communication over GSM network for the purpose of controlling remote devices. The system consists of three main components: voice-activated system, M2M system, and the device to be controlled. Detailed description of the hardware and software implementation of each design element is presented. The performance of the proposed system is evaluated through experimental tests on real spoken words. Such system design could be used in many applications such as military.

1. Introduction

Wireless M2M (machine-to-machine or mobile-to-machine) is a new technology that allows machines to transmit or receive data remotely over cellular telecommunication networks such as GSM (global system for mobile communication). M2M is seen as a potential powerful means of monitoring and controlling machines and promises to offer time and cost saving in many applications. This is done by utilizing the existing features in the GSM mobile handset such as dual tone multi-frequency (DTMF) and short message service (SMS). DTMF enables sending coded message of a few digits using ordinary speech channels for simple remote security or control of personal data, as in banking applications. SMS, on the other hand, is well suited to M2M communication because of its reliability and its ease of use. It is just writing alphabetic message that has information up to 140 bytes per SMS that can be sent in milliseconds to control or provide more detailed information about the machine.

The basic idea of M2M implementation is that: 1) a user writes a text message that is used for example to control the machine and sends via SMS message from the user's mobile handset to the M2M system; 2) upon receiving the SMS message, the M2M system (connected to the machine or instrumentation to be controlled) starts decoding, processing, and executing the message via software operation; 3) if the received message matches the pre-stored text message, the M2M system will turn the machine on, otherwise the system will reject it; and 4) the state of the machine will stay on until it receives another text message to turn it off. The challenge in the design of complete M2M system is to integrate the M2M system to the target machine to perform certain application task. Recently, researchers have proposed this modular integrated system in many applications such as: 1) in telemedicine for improving the quality of health care and reduce its cost by faster communication of medical information between physicians and patients (Woodward, 2001; Aranguren, 2002); 2) in home automation for monitoring and controlling home appliances (Alheraish, 2004); 3) in vending machine for providing information about the stock of products (DeAzevedo, 2003). The developed M2M systems in these works involve writing the text message from

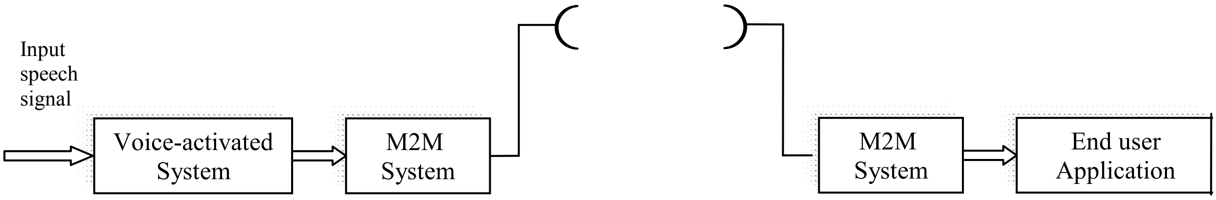

Fig. 1. Block diagram of the proposed prototype system.

mobile phone keypad. However, speech is a natural communication means for human beings and is the most familiar and convenient part of a person's capabilities, and thus it can be used effectively to control instrumentation in that person's environment (Grattan, 1991).

In this paper, a basic system design that would use voice to establish M2M communication over GSM network is developed. The use of voice for M2M communication is of an utmost importance in many applications. In military, voice-based control commands are needed when operating in hostile environment and time becomes of critical value. In addition, the developed system would help disabled users who have insufficient manual dexterity to operate mobile telephone keypad. The developed system would even help able-bodied users to do fast and quick control actions.

The proposed system consists of voice-activated system and M2M system in the sending end and M2M system and end user controlled machine in the receiving end. This is discussed in Section 2. Section 3 describes feature extraction and classification stages as main components of the voice-activated system. For feature extraction stage, we propose a simple and yet easy to implement linear predictors that are able to extract the important features for discrimination between spoken words. For classification stage, a neural network is used to train the extracted features and generate numeric values of spoken words. The M2M system presented in Section 4 is designed as M2M module and microcontroller (or PC). The software development of the complete proposed system is outlined in Section 5. To evaluate the feasibility of the proposed system over GSM network, we conduct two lab experiments in Section 6. The required electronic designed circuits are also shown in this section. Finally, the concluding remarks are reported in Section 7.

2. Proposed System Design

The block diagram of the proposed system is shown in Fig. 1. In the sending end, the speech signal is fed to a voice activated system, which will recognize the spoken words and convert them into suitable codes. The codes are then presented as text to the M2M system using software programming. The M2M system works as a mobile modem, which will send the codes as an SMS message over the GSM network. In the destination terminal, the M2M system receives the SMS message, decode the text message, and send it to the instrumentation that will be controlled. These functions of M2M system are executed by a program that processes the incoming data and implements the output on the instrumentation. The instrumentation is either electric or mechanic target a vast array of automotive applications. In order to apply an M2M to an instrumentation, there must be a way to communicate the instrumentation with the M2M system. In the case of the electrical machine, the communication will be via a direct interface, but in the case of the mechanical machine sensors, they have to be placed before the interface to convert the mechanical signals into electrical signals. The M2M system is equipped with a serial data interface fully compatible with the standard RS232 interface in both mechanical and electrical specifications. The detailed design of each subsystem in Fig. 1 and its function in the overall system design is presented next.

3. Voice-activated System

The voice activated system consists of three main blocks: analog-to-digital (A/D) converter, feature extraction, and classification as shown in Fig. 2. The speech signal is first applied to analog-to-digital converter so as to allow discrete-time processing of the continuous-time speech signal. The feature extraction stage reduces the dimensions of the input

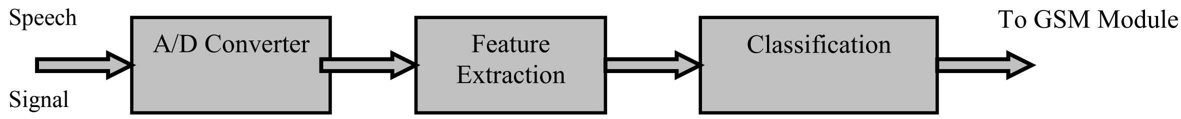

Fig. 2. Block diagram of voice activated system.

speech signal and extracts the important features for discrimination between spoken words. Finally, the classification stage makes use of the extracted features to assign an AT (Attention) command to the input speech signal drawn from a fixed number of speech patterns. The AT command is then transferred to the M2M system for establishing the intended communication link and performing the required task at the receiving end. The function of each sub-block is explained below.

3.1. Feature extraction

Most algorithms model speech data using the method of linear prediction. This method has become the predominant technique for representing speech data for low bit rate transmission or storage. The importance of this method lies both in its ability to provide extremely accurate estimates of speech data parameters, and in its relative speed of computation. The basic idea behind linear predictive modeling is that a speech data sample *x(n)* can be approximated as a linear combination of past speech data samples *x(n-1), x(n-2),* . . . etc. The best modeling advantage is obtained when the model coefficients are chosen such that the residual signal denoted by *e(n)* is minimized (e.g. in the mean-square sense) over the analysis frame during which the characteristics of the speech signal are reasonably stable. One obvious virtue of this model is that if it is accurate, we should be able to predict future values of the signal based on our current set of measurements. The error term should tell us something about the quality of our model; if the error is small, the model is accurate. Much of the arithmetic in the algorithm of linear prediction described earlier is in the computation of the model parameters. In this section, we introduce three simple predictors, each of which has only constant coefficients:

$$\begin{aligned} e_0(n) &= x(n) \\ e_1(n) &= x(n) - x(n-1) \\ e_2(n) &= x(n) - 2x(n-1) - x(n-2) \end{aligned} \tag{1}$$

These predictors have been developed based on the differentiation approach. This approach is implemented by taking the difference between adjacent signal samples. For low frequency signals, the differences between consecutive samples tend to be quite small when compared to the original signal. For high frequency signals, the reverse is true; however, speech signals are generally dominated by low frequency components. In any event, examining Eq. (1), we observe that the second predictor is, in fact, a first order differentiator, while the third predictor is a second-order differentiator. An interesting property that is worth noting is that residual signals $e_1(n)$ and $e_2(n)$ of the second and third linear predictors can be efficiently computed in the following recursive manner:

$$\begin{aligned} e_1(n) &= e_0(n) - e_0(n-1) \\ e_2(n) &= e_1(n) - e_1(n-1) \end{aligned} \tag{2}$$

This permits the entire set of prediction error signals to be computed without any multiplications, which is a prime target when data modeling is performed at the transmitting end. In our proposed system, speech signal is recorded for 0.7 seconds and at a frequency of 8 KHz. Therefore, each spoken digit is represented by 56000 points. The discrete-time signal is then passed through the three linear predictors. Statistical properties of prediction error signals that are robust and insensitive to typical errors or noise have been considered. These properties include variance and number of zero-crossings. Therefore, we have a total of six features extracted for each digit.

3.2. Classification

This stage will be implemented using neural network. A neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection (synaptic) weights, and the processing performed at computing elements or nodes (Haykin, 1994). Figure 3 shows a three-layer feed-forward neural network with six inputs and one output. Each layer of processing elements makes its independent computations on data that it receives and passes on the results to the next layer. The computations performed at each node are as follows. The individual inputs μ_1, μ_2, ..., μ_L are weighted by the parameters w_1, w_2,…, w_L and the weighted values are summed up with a bias denoted by w_0 to form the activation function input as shown in Fig. 4. Many activation functions can be used. This work employs the tan-sigmoid activation function for the first and second hidden layers and the linear activation function for the output layer. Therefore, the feed-forward neural network provides an explicit non-linear relationship between input and output data.

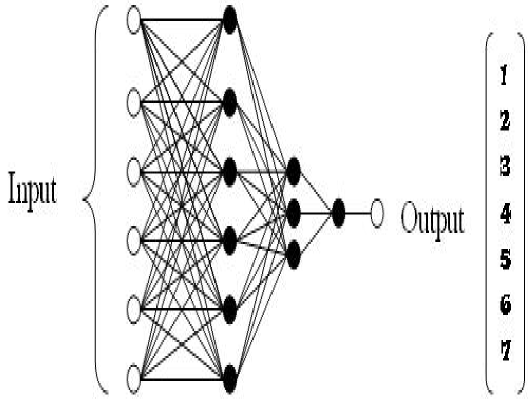

Fig. 3. A three-layer feed forward network.

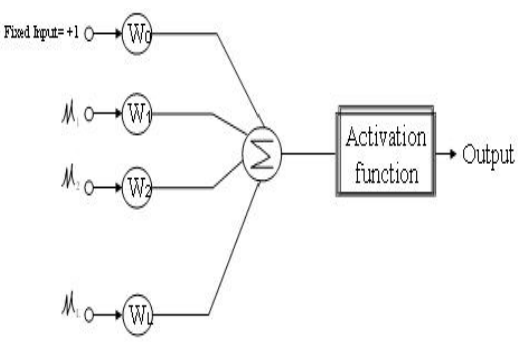

Fig. 4. A neuron with L-element input vector.

Neural networks are one of a group of intelligent methods for data analysis and modeling. They differ from other classical techniques by learning. Note that weights are the primary means of long-term storage in neural network, and learning usually takes place by updating these weights with a set of input/output pattern pairs so that the network will produce the desired output for a given input in the future.

Specifically, neural networks undergo a training session during which their free parameters (i.e. synaptic weights and biases) are adjusted in a systematic way so as to minimize a cost function. Typically, the cost function is defined on the basis of a mean squared-error criterion, with the error itself being defined as the difference between a desired response and the actual output of the network produced in response to a corresponding input signal. The neural network learns from examples by constructing an input-output mapping for the problem at hand without the need for knowing the underlying probability distribution of the data.

Assume that we have *K* patterns: x_1, x_2…, x_K with each pattern has *N* elements. We arrange these patterns in a matrix A whose *i*th row is denoted by x_i. Associated with the pattern matrix A is a desired M × K output matrix D. The i^{th} row of D is the desired output of x_i. The problem is then to determine the values of unknown network weights so that the error between the desired output and the actual output is reduced. In our development, the back-propagation algorithm (Heykin, 1994) has been used to adjust the weights of a three-layer feed-forward network with the first hidden layer has 6 nodes, the second hidden layer has 3 nodes, and the output layer has 1 node. The network is trained with a set of 250 input/output pattern pairs obtained from the recording of each digit seven times. The entries of each input pattern are the variances and number of zero-crossing of each predictor. The corresponding entry of the associated output pattern is the numeric value of the spoken digit. Figure 5 shows the difference between the actual and estimated digit at the output of the neural network when tested with 100 new patterns. It is evident from the result obtained that the network is capable of recognizing more than 90% of the input data.

4. M2M System

Basically, M2M system consists of M2M module and processor (either microcontroller or PC) as shown in Fig. 6. The M2M module (or GSM modem) works like a mobile telephone, but it does not have neither keys nor display because it has been designed to be used by a machine. The M2M module acts like an interface between the microcontroller and the GSM network. The M2M engine makes the system log on the GSM network and ready to make any communication or transferring of data. The module has a built in Subscriber Identity Module card (SIM card) to make the network identify the user and

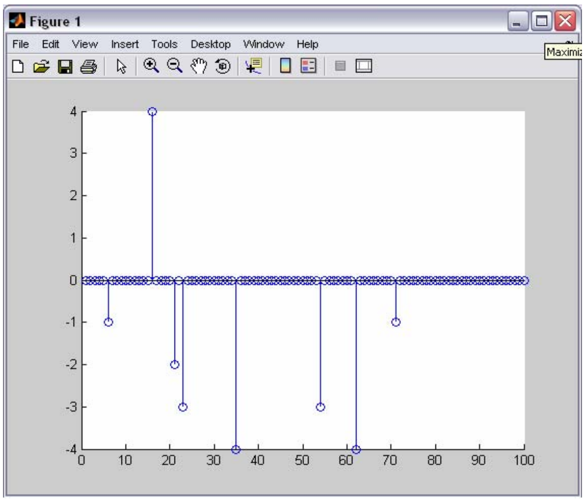

Fig. 5. The difference between the actual and estimated digit at the output of neural network.

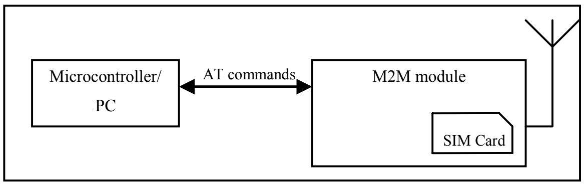

Fig. 6. M2M System.

provide the user the GSM services. The module takes the AT commands from the microcontroller and send them to the other end of M2M system by the GSM network. The AT commands have a very large number of commands; each command performs a certain task. The microcontroller (or PC) is able to communicate with the M2M module when there is a need to access the network for sending or receiving data. Since the microcontroller is an electronic device, the signals that enter it must be electrical. In order to make an electrical or mechanical machine communicate with the microcontroller, an interface has to be used. The microcontroller takes the data from the interface and makes some calculations if needed then translates the data into AT commands so the module can understand it. The controller also takes the data from the module then translates the data to the interface to control the machine.

5. Software Construction

We have implemented the prototype system design using two programming software, one for voice-activated system and the other for M2M system. The general algorithms of the two programs are shown in Fig. 7. In the voice-activated system, a MATLAB code is written to process the real-time input speech and convert it to a suitable code. The following software algorithm outlines the steps used to perform the desired task:

- Initialize the program by recording the speech segment.
- Compute the filtered version signal of the speech segment using Eq. (1).
- Compute variance and number of zero-crossings feature of each filtered version signal.
- Input the features in step 3 to neural network and do the recursive data training and testing.
- Record the output in a file.

In M2M system, the software processes the output of the voice-activated system by running a visual C++ program. At both sending and receiving ends, the PC is communicated with M2M module via C++ program. This process is outlined as follows:

- Set-up the M2M module by initializing RS232 connection and including all required library files.
- Initialize the M2M module at both sending and receiving ends.
- At the sending end:
  1. Set-up the necessary AT commands: AT+CSDH=1 to enable the transmission of short messages; AT+CSMP=1 to set-up text mode parameters; AT+CMGF=1 to select the message in text format; and AT+CMGS=1 to send the message to the intended mobile number.
  2. The software program will periodically check and read any received text messages from voice-activated system.
  3. If the received message matches any of the pre-stored messages (that are stored according to the application requirements), then it will be sent as SMS message through GSM network; otherwise it will be deleted.
- At the receiving end:
  1. The program code will periodically check and read any received message from the M2M module using AT commands. The command that performs this task is AT+CMGR=1, which will read only message No. 1 in the memory of the M2M module.
  2. If there is an incoming message, the message will be stored as a string; otherwise it will be deleted. The message is deleted using AT+CMGD=1 command.

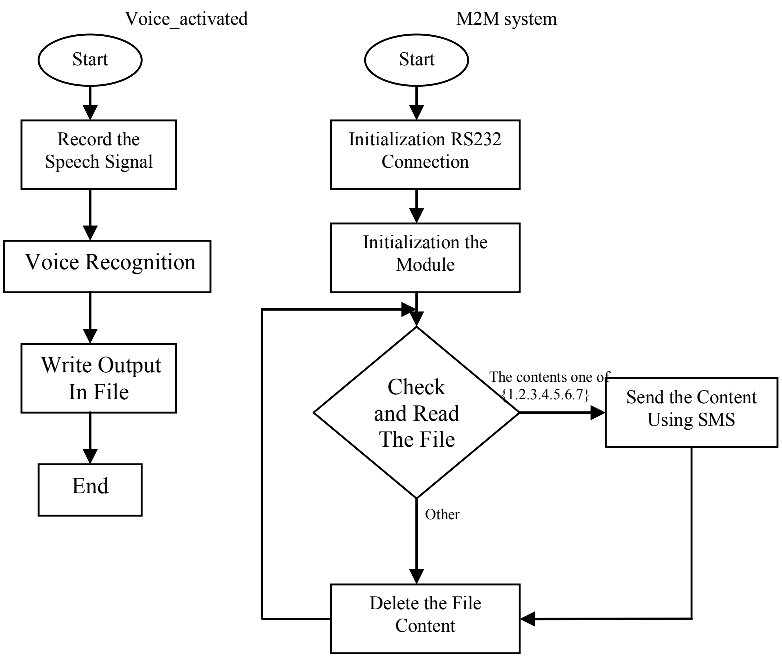

Fig. 7. Flowchart of voice-activated and M2M software implementation.

3. Extract the sending text. Since the SMS message includes details such as date, time, sender's number and the text message, the program must decode the text message and exclude all other details. If the received message matches any of the pre-stored messages (that are stored according to the application requirements), then it will be sent to the application control action; otherwise it will be deleted.

6. Development Results

This section introduces two experiments performed in the lab to evaluate the feasibility of our prototype system design of voice-activated M2M communication system over GSM network. In the experiments, we used GM47 module developed by Sony Ericsson as an M2M module (Ericsson, 2002). The GM47 module is intended for use in 900/1800 and 850/1900 MHz GSM bands respectively. The module is used to make a connection to the GSM network and send and receive SMS and GPRS services and to make voice calls as well. The GM47 module is not a stand-alone device; it is used as an engine of the M2M system. The electrical connections to the module are made through the system connector interface. The system connector is a 60-pin that may be used as input, output or both. The module has the ability to produce output voltage that can be used to operate other machines. The maximum output is 2.75 V at pin 20. The voltage across the pin varies from 0 to 2.75 V with 256 levels that give the programmer the ability to choose the proper voltage such that 0 means 0 V and 255 means 2.75 V.

The objective of the first experiment is to send a spoken digit as an SMS message and display this digit on an 8-bit digit display screen at the receiver. At the sending end, the tested speech patters are the digits one, two, and three up to seven. The spoken digit is first recognized using neural network, and then the GM47 module associates the digit with an AT command that initiates phone call. The code of this digit is transmitted as SMS message over the GSM network. At the receiving end, the content of the SMS message is displayed on the 8-bit screen by converting the code to the corresponding digit. We have written the program needed to control the GM47 module, extract the code from the short message, and send it in the appropriate format to the output application circuit as shown in Fig. 8. This circuit is connected between pin 20 (analog output) and pin 2 (ground) of the module pins. The circuit consists of eight comparators, seven resistances, priority encoder, BCD-to-seven segment, and 8-bit display. In the experiment, the seven digits are given suitable correspondent voltages: digit 1~ 0.35 V, digit 2~ 0.70 V, digit 3 ~ 1.03 V, digit 4 ~ 1.40 V, digit 5 ~ 1.73 V, digit 6 ~ 2.06 V, digit 7 ~ 2.40 V. Thus in the circuit, if the input volt is 1.4, then the four lower comparators have output, and the output of priority encoder will be "100". This code when applied to the BCD-to-seven segment will activate the lines b, c, f and g. Hence, the number four will show on the 8-bit display. Overall, the complete proposed system accurately recognized all the digits and produced the correct spoken digits on the display screen.

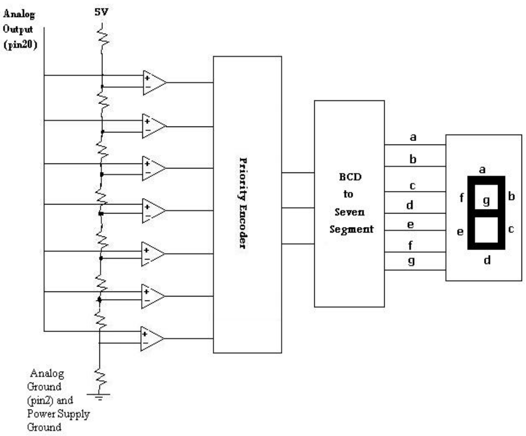

Fig. 8. Circuit connection for displaying a digit.

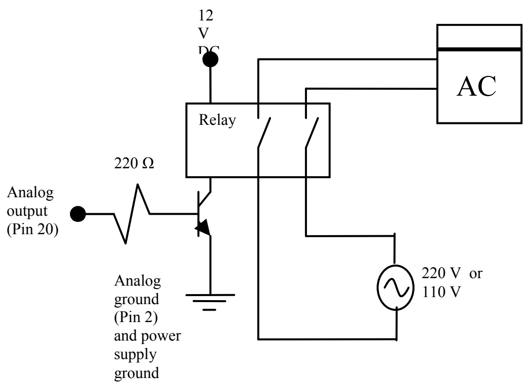

Fig. 9. Circuit connection for controlling AC device.

In the second experiment, the proposed system is tested to control air conditioner (AC) remotely. The user selects speech input of digit one for ON AC and digit two for the OFF AC. In this case, the GM47 module is only programmed to send SMS of these two digits. If the received message is not digit one or two, the system will reject it. But, if the message is digit one, the AC will turn on. The AC will not change until a message of digit two is received. The AC will not be affected if any message is received except digit two which will turn off the AC. The design circuit of this application is shown in Fig. 9. In this design, two voltages: one volt and two volt are chosen to represent digit one and digit two respectively. The programmer can also choose more voltage steps to operate the AC in low, medium and high speed levels. Since the GM47 operates with maximum voltage of 2.75 V, the circuit in Fig. 9 includes an additional circuit of the following: Transistor, 12V DC power supply, 220 Ω resistance, relay, and 220 or 110 V AC.

7. Conclusions

We have presented a novel prototype design of voice-activated M2M communication system over GSM network. We have approached the voice recognition part of this design by using a feed-forward neural network with two hidden layers having only six and three neurons, respectively. The output layer has one neuron. The features applied to the neural network are derived from the output of three simple linear predictors. The recognition system consisting of the feature extraction and classification stages, although simple in structure, has the capability to achieve recognition rate more than 90%. The complete proposed system was tested in lab to display 8-bit digit figures on the attached screen and to control air conditioner (AC) remotely. In both tests, the proposed system was able to recognize the spoken words correctly and reliably.

Future directions of the current work include the development of more robust techniques for salient feature extraction, improvement of classification stage performance, replacing the PC with a microcontroller to make the M2M system operates as a stand-alone system, as well as reducing the complexity of the sending end by eliminating the M2M system and relying only on the M2M system in the receiving end that may performed better than the one proposed in this work.

References

Alheraish, A. "Design and Implementation of Home Automation System." *IEEE Transactions on Consumer Electronics*, Vol. 50, No. 4, (Nov., 2004), 1087-1092.

Aranguren, G.; Nozal, L.; Blazquez, A. and Arias, J. "Remote Control of Sensors and Actuators by GSM." *IEEE 2002 28th*

Annual Conference of the Industrial Electronics Society IECON 02, Porto, Portugal, Vol. 3, (5-8 Nov., 2002), 2306-2310.

DeAzevedo, D.; DeMoura, E.; DeCastro, M. and Da Rocha, F. "Telemedicine: Remote Monitoring of Cardiac Patients." *25th Annual IEEE EMBS International Conference*, Cancun, Mexico, (Sept., 2003).

Ericsson Telecom Environment. *GM47/GM48 Developer's Kit.* Sony Ericsson Mobile Communication International, (2002).

Grattan, K.; Palmer, A. and Shurrock, C. "Speech Recognition for the Disabled." *IEEE Magazine Engineering in Medicine and Biology*, Vol. 10, No. 3, (Sept., 1991), 51-57.

Heykin, S. *Neural Networks*. New York, NY: Macmillan College Publishing Company, (1994).

Woodward, B.; Istepamian, H. and Richards, C. "Design of a Telemedicine System Using a Mobile Telephone." *IEEE Transaction on Information Technology in Biomedicine*, Vol. 5, No. 1, (March, 2001), 13-15.

التحكم بالأجهزة عن ُبعد باستخدام
الصوت ء شبكة ا وال

عبدا سن ا ريش، وصا الشبيلي، وفهد الشهري، و مد أبو العلا

قسم الهندسة الكهربائية ، كلية الهندسة ، جامعة الملك سعود ، ص ب ٨٠٠ ،
الرياض ١١٤٢١ ، المملكة العربية السعودية

(قدم للنشر في ٢٢/١١/٢٠٠٦م ؛ وقبل للنشر في ٢٨/٠٤/٢٠٠٧م)

ملخص البحث. تعتبر تقنية التحكم بالأجهزة (M2M) من التقنيات الحديثة التي تتيح إرسال المعلومات عبر شبكة الجوال. يهدف هذا البحث إلى تصميم نظام جديد يستخدم الصوت في مجال التحكم عبر قناة الجوال لخدمة المعاقين ومستخدمي الجوال في ظروف حرجة (على سبيل المثال أثناء قيادة السيارة)، وتتلخص فكرة هذا التصميم في اختيار عدة كلمات صوتية يتم تحويلها من خلال برنامج حاسوبي في جهاز الجوال المرسل له رسائل قصيرة تحمل كل رسالة أمر يمكن تنفيذه من خلال جهاز تحكم مرتبط بالجوال المستقبل للرسالة. ويعتمد البرنامج الحاسوبي على استخدام الشبكات العصبية الاصطناعية للتعرف على الكلمات الصوتية ومن ثم إرسال الأمر المقابل لكل كلمة. ولتقييم أداء هذا التصميم، عرض البحث عددًا من التجارب المعملية باستخدام أصوات حقيقية.