Maximum Torque per Ampere Control of Brushless Doubly Fed Induction Generator Using Variable Structure Approach for Wind Turbine Applications
Hamid Reza Mosaddegh
Hossein Aboororabi Zarchi

Analysis on Radio-Frequency Modeling of Double- and Single-Gate Square-Shaped Extended Source TFETs
Saeid Marjani
Seyed Ebrahim Hosseini

A Measurement Setup for Radiated EMI of Un-buffered DRAM Modules
Mojtaba Joodaki
Amir Attar

Incorporation of Distribution System Reconfiguration and Expansion Planning Problems by Considering the Role of Demand Response Resources
Hamidreza Arasteh
Mohammad sadegh Sepasian
Vahid Vahidinasab

Study on Health Monitoring of Concrete Structures Using Wireless Sensor Networks
Saman Shooreh Sani
Majid Baghaei-nejad
Mona Kalate Arabi

A Novel Approach to Speaker Weight Estimation Using a Fusion of the i-vector and NFA Frameworks
Amir Hossein Poorjam
Mohamad Hasan Bahari
Hugo Van hamme
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Torque per Ampere Control of Brushless Doubly Fed Induction Generator Using Variable Structure Approach for Wind Turbine Applications</td>
<td>Hamid Reza Mosaddegh, Hossein Abootorabi Zarchi</td>
<td>1</td>
</tr>
<tr>
<td>Analysis on Radio-Frequency Modeling of Double- and Single-Gate Square-Shaped Extended Source TFETs</td>
<td>Saeid Marjani, Seyed Ebrahim Hosseini</td>
<td>9</td>
</tr>
<tr>
<td>A Measurement Setup for Radiated EMI of Un-buffered DRAM Modules</td>
<td>Mojtaba Joodaki, Amir Attar</td>
<td>15</td>
</tr>
<tr>
<td>Incorporation of Distribution System Reconfiguration and Expansion Planning Problems by Considering the Role of Demand Response Resources</td>
<td>Hamidreza Arasteh, Mohammad Sadegh Sepasian, Vahid Vahidinasab</td>
<td>23</td>
</tr>
<tr>
<td>Study on Health Monitoring of Concrete Structures Using Wireless Sensor Networks</td>
<td>Saman Shoorabi Sani, Majid Baghaci-Nejad, Mona Kalate Arabi</td>
<td>37</td>
</tr>
<tr>
<td>A Novel Approach to Speaker Weight Estimation Using a Fusion of the i-vector and NFA Frameworks</td>
<td>Amir Hossein Poorjam, Mohamad Hasan Bahari, Hugo Van hamme</td>
<td>47</td>
</tr>
</tbody>
</table>
Maximum Torque per Ampere Control of Brushless Doubly Fed Induction Generator Using Variable Structure Approach for Wind Turbine Applications

Hamid Reza Mosaddegh and Hossein Abootorabi Zarchi

Abstract. In this paper, a variable structure controller is proposed to control the torque of brushless doubly fed induction generator with considering maximum torque per Ampere (MTPA) strategy. Based on this control approach, a sliding mode controller with a PI is designed in order to reduce the torque pulsations during steady-state behavior while the fast response and robustness merits of the classic DTC are preserved. Also this method decreases the copper losses with minimizing the magnitude of power and control windings currents without deteriorating the dynamic performance. The presented method, guarantees maximum power point tracking (MPPT) with a desirable operation. The simulation results verify an accurate, quick and robust operation of the brushless doubly fed induction generator for wind turbine applications.

Keywords: Brushless doubly fed induction generator (BDFIG), Maximum Torque Per Ampere (MTPA), Sliding mode control (SMC), Torque control, Wind turbine.

1. Introduction

Brushless doubly fed induction generator (BDFIG) is a single frame and brushless machine which has two 3-phase windings that are set on the stator. One of them is power winding (PW) that is directly connected to the grid. Most of power is exchanged between the BDFIG and grid through this winding. The other winding connected to the grid through a back-to-back converter with less capacity than the generator, is called control winding (CW) [1]. The converter at generator side has the duty of speed control and controlling the reactive power of generator but the converter at grid side, controls the voltage of DC link and regulates the terminal voltage by absorbing or supplying reactive power [2]. It must be mentioned that the control winding capacity is dependent on the desired speed range and reactive power requirements [3].

Because of some advantages like elimination of brushes (high reliability and low maintenance) and the use of a fractionally-rated frequency converter, BDFIG have been paid attention to use in some industrial operations like wind – power generation [1].

The special rotor structure of this generator increases the complexity and produces undesirable spatial harmonics which decrease machine efficiency [4].

Due to limitations of efficiency improvement through the machine design, in recent years, several publications have been released in the field of ac machine drives optimization techniques [5]-[7], but few papers have been reported regarding access to maximum efficiency for BDFIG.

For high performance drives, it is desirable to achieve optimum operation at MTPA. MTPA strategy is a smart answer to the call for efficiency. In principle, the target of MTPA strategy is to deliver the electromagnetic torque with the lowest current magnitude. In this way, copper losses are minimized and the overall system efficiency is increased, at least as long as copper losses are dominant [8]. It should be mentioned that the core losses in a BDFIG are not offered yet, so it is impossible to formulate all loss components and analyzed them, numerically [9]. Therefore, in this paper, core losses are neglected while the stator currents are minimized under the constraint of constant torque at a certain speed.

In this research, a nonlinear controller is introduced on the basis of Linear and Variable Structure Control (LVSC), to drive the BDFIG in wind turbine systems, considering the MTPA control strategy. The proposed method improves the operation of control system by capturing the maximum torque from wind and hence increases the efficiency of wind turbine. Also, without deteriorating the dynamic performance, the presented MTPA approach decreases the copper losses with minimizing the magnitude of power and control winding currents.

2. Introduction to BDFIG

A. BDFIG Operation

The stator of this generator has two windings which have different numbers of pole pairs to prevent direct coupling between them. Also, in order to reduce the electromagnetic forces on the rotor, the difference between pole pairs must be greater than one [10]

\[|P_1 - P_2| > 1,\]  

where \(P_1\) and \(P_2\) are PW and CW pole pairs, respectively.

The rotor of this generator is designed with a special manner. The most conventional structure used in the rotor of this generator is nested-loop. The number of nests which is the number of rotor poles is equal to the sum of PW and CW pole pairs so as to cause indirect coupling between CW and PW [2].
Owing to special structure of rotor, there will be different modes of operation for this machine. But the best operation of BDFIG is achieved in synchronous mode. In this mode, the frequency of induced voltage in PW, due to indirect coupling with CW, is equal to grid frequency. This situation leads to generation of two fields that turn at the rotor speed. Also, according to the number of rotor poles, in order to establish the indirect coupling between PW and CW, that is fundamental of torque generation in this machine, the direction of rotation of PW magneto-motive force (mmf) respect to the rotor, will be in opposite direction of CW mmf. In this condition \[10\]

\[
\omega_1 - p_1 \omega_r = -(\omega_2 - p_2 \omega_r)
\]  
(2)

So, the synchronous rotor speed is determined as follow:

\[
\omega_r = \frac{\omega_1 + \omega_2}{p_1 + p_2}
\]  
(3)

where \(\omega_1\) and \(\omega_2\) are PW and CW angular speed, respectively.

If the CW current is dc, the natural speed will be achieved as \[11\]

\[
\omega_n = \frac{\omega_1}{p_1 + p_2}
\]  
(4)

The slips for the two windings are defined as \[11\]

\[
s_1 = \frac{\omega_1 - \omega_r}{p_1} = \frac{\omega_1 - p_1 \omega_r}{\omega_1}\n\]  
(5)

\[
s_2 = \frac{\omega_2 - \omega_r}{p_2} = \frac{\omega_2 - p_2 \omega_r}{\omega_2}\n\]  
(6)

where \(s_1\) and \(s_2\) are PW and CW slips, respectively.

The BDFIG can be modeled as two induction generators in a common frame. The equivalent circuit of machine is depicted in Fig. 1. The parameters of mentioned circuit are referred to PW side and the power losses are neglected. This circuit is valid for all operation modes of BDFIG including synchronous mode \[12\].

**B. BDFIG Model**

The generator model in PW flux frame is expressed by the following \[1\]

\[
V_1 = R_i I_1 + j \omega_r I_1 + j \omega_1 \psi_1
\]  
(7)

\[
\psi_1 = L_1 I_1 + L_{1r} I_r'
\]  
(8)

\[
V_2^* = R_i^* I_2^* + \frac{d\psi_2^*}{dt} + j(\omega_1 - (p_1 + p_2) \omega_r)\psi_2^*
\]  
(9)

\[
\psi_2^* = L_2^* I_2^* + L_{2r} I_r'
\]  
(10)

\[
V_r' = R_{r'} I_r' + \frac{d\psi_r'}{dt} + j(\omega_1 - p_1 \omega_r)\psi_r'
\]  
(11)

\[
\psi_r' = L_{r'} I_r' + L_{2r} I_1 + L_{2r}^* I_2^*
\]  
(12)

In the above equations, the subscripts 1, 2, and \(r\) show PW, CW, and rotor, respectively.

The electromagnetic torque is given by \[9\]

\[
T_e = \frac{1}{2} \left[ \left( \frac{n_1}{n_2} \right)^2 L_{2r}^2 I_2^2 - I_1^2 I_r' \right] R_{r'}^2 \frac{\omega_1}{Z_r}
\]  
(13)

\[
- \frac{n_1}{n_2} I_1^2 I_{1r}^2 I_2^2 \frac{\omega_1 Z_r}{Z_r} \left[ (p_2 - p_1) \cos(\psi) \cos(2\delta) \right. 
\]  
\[
+ (p_2 + p_1) \sin(\psi) \sin(2\delta) \right],
\]  
(13)

where

\[
l_1 = I_2^* = l \Rightarrow l_2 = \left( \frac{n_1}{n_2} \right) l
\]  
(14)

\[
\delta = \tan^{-1} \left( \frac{i_q}{i_d} \right)
\]  
(15)

\[
\omega_s = p_1 \omega_r - \omega_1
\]  
(16)

\[
\psi = \tan^{-1} \left( \frac{\omega_s L_2'}{R_r'} \right)
\]  
(17)

\[
Z_r = \sqrt{R_{r'}^2 + (\omega_s L_1')^2}
\]  
(18)

Also, the mechanical equation of generator is

\[
T_L - T_e = B \omega_r + J \frac{d\omega_r}{dt}
\]  
(19)

where \(J\) is the moment of inertia, \(B\) is the friction coefficient and \(T_L\) is the torque produced by wind turbine. The parameters and variables used in (7) – (12) are introduced in Table 1.
### Table 1. BDFIG parameters and variables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V, I, \psi, \omega_r$</td>
<td>Voltage, current, flux vectors and generator speed</td>
</tr>
<tr>
<td>$R_1, R_2, R_r$</td>
<td>Resistances of PW, CW and rotor</td>
</tr>
<tr>
<td>$L_1, L_2, L_r$</td>
<td>Self-inductances of stator windings and rotor</td>
</tr>
<tr>
<td>$L_{1r}, L_{2r}$</td>
<td>Coupling inductances between stator windings and rotor</td>
</tr>
</tbody>
</table>

### 3. Control Strategy

#### A. MTPA control

The MTPA strategy is obtained with minimization of CW and PW current magnitude under the constraint of constant torque at a certain speed. Based on Lagrange theorem, it can be easily found that any control strategy can be realized when the torque and its corresponding objective function are tangent at a point or in other words their gradient vectors are in parallel. In MTPA, the minimization of stator current is selected as objective function.

The square PW current of BDFIG can be calculated as

$$I_1^2 = I_{1d}^2 + I_{1q}^2. \quad (20)$$

In this section, the minimization of (20) under the constraint of constant torque is selected as objective function. In Fig. 2, according to (13), the constant torque curve can be drawn as a hyperbola on the $I_{1d} - I_{1q}$ plane.

On the same plane, at a speed and with neglecting the iron losses, the curve that representing the square PW current, takes the form of a circle. Under the constraint of constant torque, if an operating point is set at point “a” in Fig. 2, the curve A is supposed to be a control objective curve such as constant stator current curve ($I_1^2$). If an operating point is set at “b”, the curve B is another control objective curve ($I_1^2$). Using Lagrange’s Theorem, it can be easily found that control objective is minimum when the torque curve and control objective curve are tangent at a point if and only if their gradient vectors are in parallel. This means that $\nabla T_e(I_{1d}, I_{1q})$ must be a scalar multiple of $\nabla I_1^2(I_{1d}, I_{1q})$ at the point of tangency (see “b” in Fig. 2), so that

$$\nabla T_e(I_{1d}, I_{1q}) \parallel \nabla I_1^2(I_{1d}, I_{1q}) \sin \theta = 0, \quad (21)$$

where $\theta$ is the angle between $\nabla T_e(I_{1d}, I_{1q})$ and $\nabla I_1^2(I_{1d}, I_{1q})$.

Criterion of MTPA strategy realization is obtained as follows:

$$y_1 = \nabla T_e(I_{1d}, I_{1q}) \cdot \nabla I_1^2(I_{1d}, I_{1q}) \sin \theta. \quad (22)$$

It is obvious that the control strategy is realized when $y_1$ is kept at zero. The cross-product of gradient vectors is calculated from the following equation

$$\nabla T_e(I_{1d}, I_{1q}) \times \nabla I_1^2(I_{1d}, I_{1q}) = \det \begin{bmatrix} \frac{\partial T_e}{\partial I_{1d}} & \frac{\partial T_e}{\partial I_{1q}} & 0 \\ \frac{\partial^2 I_1}{\partial I_{1d}^2} & \frac{\partial^2 I_1}{\partial I_{1q}^2} & 0 \\ \frac{\partial I_1}{\partial I_{1d}} & \frac{\partial I_1}{\partial I_{1q}} & 0 \end{bmatrix}, \quad (23)$$

So $y_1$ will be obtained as follows

$$y_1 = -I_{1q} \left( \frac{1}{2} \tan^{-1} \left( \frac{p_2 + p_1}{2} \right) \sin \psi \right) \quad (24)$$

With neglecting the core losses, criterion of realization of the MTPA strategy in BDFIG is obtained as (24). Although ideally, the optimum current angle is constant ($\delta \approx 45^\circ$), however, with considering core losses, value of this angle is greater than ideal amount.

#### B. Linear and Variable Structure Control

Due to nonlinear nature of electrical machines, if the reference voltage is generated by a nonlinear controller, a better operation of electrical drive will be achieved. In this regard, the sliding mode controller because of robustness to uncertainties and variations in system parameters, fast dynamic response and also compensation of disturbance effects, have been paid attention by electrical drive researchers. However, robustness of this controller is only in its sliding phase and the reaching phase is designed so that the mechanical mode paths of system reach to sliding phase as quick as possible. In other words, the dynamic of system is not perfectly robust all the time. So, it is possible that the conventional SMC can not retain its stability against the uncertainties and disturbances. Also the other drawback of this controller is chattering effect, which might be harmful for the system.
To improve the stability of system against uncertainties and also elimination of chattering in control system, in this paper a control method which is a combination of linear controller (PI) and SMC, introduced as Variable Structure Control (VSC), is proposed. This method, while having the simplicity of implementation, has the greatest features of linear controller (i.e. smooth and without chattering operation) and the feature of SMC (i.e. robustness against uncertainties). Another advantage of this method is independence to system parameters.

The block diagram of Linear and Variable Structure Control (LVSC) that is implemented on generator side converter is shown in Fig. 3. The main task of LVSC is fast and reliable access to control of torque and also achieving the MTPA. An SVM unit is used to produce switching signals based on voltage references.

The sliding surface is defined as below

\[ S_1 = e_{y_1} + c_{y_1} \frac{de_{y_1}}{dt} \]  
\[ S_2 = e_{y_2} + c_{y_2} \frac{de_{y_2}}{dt} \]

where \( e_{y_1} = y_1^* - \hat{y}_1 \) is error between the reference \( y_1^* = 0 \) and the cross product of \( \nabla T_e(I_{1d}, I_{1q}) \) and \( \nabla I_1(I_{1d}, I_{1q}) \). In addition, \( e_{y_2} = y_2^* - \hat{y}_2 \) is torque error.

Superscript "*" shows measured values and "\( \hat{\} \)" shows reference values. Design constants \( c_{y_1} \) and \( c_{y_2} \) are selected so as to impose the desired dynamics in sliding mode. CW reference voltage, \( V_2^* = V_{2d}^* + jV_{2q}^* \) is gained at the output of variable structure controller that \( V_{2d}^* \) is obtained by the MTPA control and \( V_{2q}^* \) is obtained by torque control law. Sign function, \( Sgn(\cdot) \), is used in control law.

\[ V_{2d}^* = (K_{y_1} + \frac{K_{y_1}}{s})(e_{y_1} + K_{VSCy_1} Sgn(S_{y_1})) \]  
\[ V_{2q}^* = (K_{y_2} + \frac{K_{y_2}}{s})(e_{y_2} + K_{VSCy_2} Sgn(S_{y_2})) \]

In the above equations, \( s \) is Laplace operator, \( K_{y_1} \), \( K_{y_1} \), \( K_{y_2} \), \( K_{VSCy_1} \), \( K_{VSCy_2} \) are PI controller gains and \( K_{VSCy_1} \), \( K_{VSCy_2} \) are variable structure control gains.

With selection of proper coefficients for linear controller and SMC, the best response from the aspect of system robustness and best time response is attainable without chattering effect. It must be considered that in transient condition, linear controller is more dominant and PI coefficients must be set so as to obtain the desired dynamic response. In steady state, SMC is more dominant and the best steady state response can be achieved by optimally setting the \( K_{VSCy_1} \) and \( K_{VSCy_2} \). It can be proved that large enough values for \( K_{VSC} \) fulfill the reaching and stability condition \( s, \dot{s} < 0 \) [14].

4. Wind Turbine and MPPT

A. Aerodynamic Model

The wind turbine extracts the energy of its blades and transmits this energy to the generator. The power absorbed from wind turbine is given below [15]

\[ P_{mech} = \frac{1}{2} \pi C_p(\beta, \lambda) \rho R^2 v^3 \]

where \( \rho \) is air density, \( R \) is turbine radius, \( v \) is wind speed and \( C_p \) is the turbine power coefficient which its maximum value is 0.59. In a wind turbine, this coefficient
is usually between 0.25 and 0.45. This quantity is depended on tip speed ratio \((\lambda)\) and the blade pitch angle \((\beta)\). \(\lambda\) is obtained by the following equation

\[
\lambda = \frac{R}{v} \omega,
\]

(30)

where \(\omega\) is the speed of turbine rotor. In this paper, \(\beta\) is supposed to be zero and so \(C_{p}\) is only depended to \(\lambda\).

B. MPPT Algorithm

The extracted power from wind turbine is depended on the accuracy of the algorithm of MPPT that ensures maximum energy yielding. The strategies can be classified into two groups [17]: the look-up table based strategies; and the strategies that are independent of aerodynamic characteristics. Some of the control algorithms that require the aerodynamic information are tip speed ratio (TSR) control, power signal feedback (PSF) control and optimum torque (OT) control. The MPPT techniques which do not need aerodynamic characteristics are often known as HCS methods. The two significant control strategies of this category include perturbation and observation (P&O) control and fuzzy logic control (FLC).

In this paper, regarding to BDFIG torque control, the best method to achieve MPPT, is the optimum torque control method. This method is superior to others because of its simplicity and accuracy. The other advantage of this method is that there is no need to have speed controller and wind speed sensor. In this technique the reference value for torque can be obtained using rotor speed and the optimum torque versus generator speed curve. In (31), the optimum torque is expressed as a function of generator speed [16]

\[
T_{opt}(\omega) = \frac{0.5\rho C_{pmax} R^5}{\lambda_{opt}^3} \omega^2 = K_{opt} \omega^2,
\]

(31)

With controlling the generator torque in \(T_{opt}\), the maximum power can be extracted from wind turbine. \(T_{opt}\) is used as the reference torque.

5. Simulation Results

In this section, the simulation results are presented to confirm the desirable operation of the proposed control method. Table 2 shows the wind turbine and generator parameters. The proper selection of controllers parameters is important to obtain the desired performance of system. In this paper, the this parameters is selected with trial and error method. The values of PI and SMC parameters are shown in Table 3.

The reference torque signal is generated by OT MPPT algorithm. As shown in Fig. 5, the generator torque tracks the reference value in variable wind speed pattern correctly. The wind speed profile can be seen in Fig. 4.

The power coefficient of the wind turbine is shown in Fig. 6. One can see that this parameter converges to its maximum value by using the proposed control system. In order to verify the realization of the presented MPPT algorithm, the desirable performance of the control system for tracking \(\lambda_{opt}\) has been illustrated in Fig. 7. It tends to capture maximum power from wind turbine and proper implementation of the MPPT algorithm.

Also, with the proposed variable structure control, the MTPA strategy realization is satisfied. The variable \(y_1\) always oscillates around its reference value \((y_{1,ref} = 0)\) that means the strategy has been achieved. The two-axis torque producing currents are shown in Fig. 10 and 11. With strategy realization, q- and d-axis currents will be equal as well, without considering core losses of this machine.
Fig. 5. Generator torque response to wind speed variations.

Fig. 6. Power coefficient variation.

Fig. 7. Maximum power point tracking performance.

Fig. 8. Generator speed variation

Fig. 9. $y_1$ & $y_{1ref}$

Fig. 10. PW $d$- and $q$-axis currents.

Fig. 11: CW $d$- and $q$-axis currents.
6. Conclusion

In this paper a variable structure torque control of BDFIG for wind turbine application was proposed and evaluated. The control strategy combines sliding mode controller and PI principles to achieve simple and robust high performance behavior. In particular, SMC makes the drive robust to disturbances and operating point variations, PI controller guarantees a smooth and without chattering operation and SVM improves the steady state responses of torque with reducing the ripple. During transient operation, the proposed approach verified good dynamic response and strong robustness to fluctuations of wind profile. This simple, quick, accurate and robust control strategy guaranties proper performance of MPPT algorithm and therefore best energy capturing from wind turbine. In addition, based on proposed control approach, the MTPA control strategy of BDFIG has been evaluated for wind power applications.

References


Hamid Reza Mosaddegh was born in Mashhad, Iran, in 1987. He received the B.Sc. and M.Sc. degrees in electrical engineering from Ferdowsi University of Mashhad (FUM), Mashhad, Iran, in 2011 and 2014, respectively. He is currently working toward the Ph.D. degree at FUM. His research interests are variable-speed ac drives, power electronics and renewable energies.

Hossein Abootorabi Zarchi received Ph.D. degree from Isfahan University of Technology, Isfahan, Iran. He was a Visiting Ph.D. Student with the Control and Automation Group, Denmark Technical University, Denmark, from May 2009 to February 2010. He is currently an Assistant Professor in the Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran. His research interests are renewable energies, electrical machines and applied nonlinear control in electrical drives.
Analysis on Radio-Frequency Modeling of Double- and Single-Gate Square-Shaped Extended Source TFETs

Saeid Marjani and Seyed Ebrahim Hosseini

Abstract. In this paper, the radio-frequency (RF) performances and small-signal parameters of double-gate (DG) square-shaped extended source tunneling field-effect transistors (TFETs) are investigated and compared with those of single-gate (SG) square-shaped extended source TFETs in terms of their cut-off and maximum oscillation frequencies and small-signal parameters. By using of a nonquasi-static (NQS) radio-frequency model, the small-signal parameters have been extracted. The results show that the DG square-shaped extended source TFET has higher transconductance, cut-off and maximum oscillation frequencies than single gate structure. The modeled Y-parameters are in close agreement with the extracted parameters for high frequency range up to the cut-off frequency. Results suggest that the DG square-shaped extended source TFETs seem to be the most optimal ones to replace MOSFET for ultralow power applications and RF devices.

Keywords: Double-gate (DG), radio-frequency (RF), nonquasi-static (NQS), extended source, tunneling field-effect transistor (TFET).

1. Introduction

Over the last decades, the tunnel field effect transistors (TFETs) with gate controlled band-to-band tunneling (BTBT) are investigated as a promising candidate to replace conventional metal-oxide-semiconductor field effect transistors (MOSFETs) for low-standby power (LSTP) and high-frequency applications. The unique and advantageous features of TFET are steep sub-threshold slope (SS), low power consumption and a very low leakage current [1]-[5]. Several works have addressed recently the devices with SS below 60 mV/decade at room temperature, both theoretically and experimentally [6]-[8]. However, all of them exhibit a minimized sub-threshold swing with a low OFF-current; their low ON-current was problematic. Consequently, it is necessary to improve the ON-current. There is a dramatic increase in the number of publications discussing the various designs to improve the ON-current by means of using band-gap engineering [9]-[15], small band-gap materials [16], [17], high-k dielectric materials [18], pocket doping [19]-[21], line-edge roughness [22], [23], vertical direction tunneling [24] and extended source [25].

Although there have been reports on the design and optimization of square-shaped extended source structure for better performance in terms of sub-threshold swing and drive current due to its enhanced tunneling area and total band-to-band generation [25], [26], their RF characterization and modeling have been seldom reported.

The goal of this paper is to determine the RF performances and small-signal parameters of double-gate (DG) square-shaped extended source TFETs and compared with single-gate (SG) square-shaped extended source TFETs. By using of the analytical equations for the Y-parameters of a nonquasi-static (NQS) radio-frequency model, the small-signal parameters were extracted for analysis of cut-off frequency, maximum oscillation frequency, gate-source capacitance, gate-drain capacitance and transconductance. This paper is organized as follows. After discussing in Section 2 the device structures and radio-frequency model of extended source TFETs, we present in Section 3 comparisons of RF performances and small-signal parameters of devices reported in the literature. We summarize our findings in Section 4.

2. Device Structure and Radio-Frequency Model of Extended Source TFETs

The cross-sectional views of double- and single-gate square-shaped extended source TFET devices used in the two-dimensional device simulation [27] are shown in Fig. 1(a) and (b), respectively. The p+ source and n+ drain are doped at 10^{20} cm^{-3}. The p-type body doping is 10^{15} cm^{-3}. The gate oxide thickness, channel length and extended source length are 2, 30 and 10 nm, respectively. By using of a dynamic nonlocal path tunneling approach, band-to-band tunneling has been modeled. For higher accuracy band-gap narrowing, Shockley-Read-Hall recombination, mobility, Auger recombination and trap-assisted tunneling models are also activated [28]-[30].

Fig. 2 shows the transistor nonquasi-static model used to extract the small-signal parameters of the square-shaped extended source TFETs in this paper. R_g, C_{gs}, C_{gd}, C_{ss}, \tau, g_{m}, and g_{ds} are the effective gate resistance, gate-source capacitance, gate-drain capacitance, source-drain capacitance, the charge transport delay, transconductance and source-drain conductance, respectively. The Y-parameters of the NQS model for the low-frequency region can be expressed as follows [31]:

\[ Y_{11} = \frac{\omega^2 R_g (C_{gs} + C_{gd})^2 + j \omega (C_{gs} + C_{gd})}{1 + \omega^2 R_g^2 (C_{gs} + C_{gd})^2} \] (1)
The small-signal parameters governing the RF behaviors can be extracted by using real and imaginary parts of equations (5)-(8) and the $Y$-parameters from the simulation results. The parameters of $g_m$, $g_{ds}$, $R_g$, $C_{gd}$, $C_{gs}$, $C_{sd}$, and $\tau$ are expressed as follows [31]:

\[
\begin{align*}
    g_m &= \text{Re}[Y_{11}]igg|_{\omega^2 = 0} \\
    g_{ds} &= \text{Re}[Y_{22}]igg|_{\omega^2 = 0} \\
    R_g &= \frac{\text{Re}[Y_{11}]}{\text{Im}[Y_{11}]} \\
    C_{gs} &= \frac{\text{Im}[Y_{11}] + \text{Im}[Y_{12}]}{\omega} \\
    C_{gd} &= -\frac{\text{Im}[Y_{12}]}{\omega} \\
    \tau &= -\frac{\omega^2 R_g C_{gd}^2}{C_{gd} + g_m R_g (C_{gs} + C_{gd})} \\
    C_{sd} &= -\frac{\text{Im}[Y_{22}]}{\omega} - g_m R_g C_{gd}
\end{align*}
\]
The cut-off and maximum oscillation frequencies of TFETs have been obtained from the high-frequency current gain and unilateral power gain data of the TCAD simulation, respectively. Extracted small-signal parameters from TCAD simulation have been used for the evaluation of the RF performances.

3. Results and Discussion

Fig. 3 shows the gate capacitance values of the DG and SG square-shaped extended TFET as a function of overdrive voltage ($V_{OV}$). Due to the formation of inversion layer of a TFET from the drain side toward the source side with increasing of $V_{OV}$, the regnant component in the capacitance between the gate and the inversion layer is gate-drain capacitance [32]. Increase in the $V_{OV}$ causes gate-drain capacitance extremely increases which is proportional to gate length, as can be confirmed in Fig. 3(a). The gate-drain and gate-source capacitance values of DG square-shaped extended TFET are larger than SG square-shaped extended TFET which should be mainly due to the much larger coupling between the gate and the source. Due to the extension of the inversion layer from the drain side toward the source side and fewer coupling between the gate and the source, the gate-source capacitance monotonically decreases, as shown in Fig. 3(b).

Fig. 4 shows the transconductance ($g_m$) of DG and SG square-shaped extended TFETs as a function of overdrive voltage ($V_{OV}$). As can be seen, DG square-shaped extended TFETs have 350 times higher transconductance than SG structure at high $V_{OV}$. It should be mainly due to the higher total field in between the source and channel that leads to an increase in the higher on-current and transconductance.

The RF figures of merit for extended source TFETs are analyzed in terms of cut-off frequency ($f_T$) and maximum frequency of oscillation ($f_{max}$). The values of the cut-off and maximum oscillation frequencies have been extracted by high-frequency current gain and unilateral power gain using the TCAD-simulated Y-parameter data, respectively. Fig. 5 compares the $f_T$ values of the DG and SG square-shaped extended TFETs as a function of $V_{OV}$. Generally, the cut-off frequency depends on the $g_m$, $C_{gd}$ and $C_{gs}$ ($f_T = \frac{g_m}{2\pi \times (C_{gd} \cdot C_{gs})}$). Since $C_{gd}$ (sum of the $C_{gd}$ and $C_{gd}$) and $g_m$ increase monotonically with the increase of $V_{OV}$ as shown in Figs. 3 and 4, $f_T$ of TFETs have the rising tendency as a function of $V_{OV}$. However the DG square-shaped extended TFETs have 350 times higher $g_m$ than SG structures at high $V_{OV}$, $f_T$ of DG square-shaped extended TFETs has improved only 150 times higher than that of SG structures. It is because the $C_{gg}$ of DG square-shaped extended TFETs at high $V_{OV}$ is about 2.3 times higher than that of SG structures.

Fig. 6 shows the $f_{max}$ values of the DG and SG square-shaped extended TFETs obtained from unilateral power gains as a function of $V_{OV}$. DG square-shaped extended TFETs have higher $f_{max}$ values than SG square-shaped extended TFETs because of lower channel resistance and higher $f_T$ and $g_m$. The maximum $f_{max}$ values of the DG and SG square-shaped extended TFETs were about 37.6, and 3.5 GHz at $V_{DS} = 0.7$ V, respectively. The results indicate
that DG square-shaped extended TFETs can have higher cut-off and maximum oscillation frequencies for high-frequency electronics applications.

The modeled $Y$-parameters by SPICE simulation were compared with the values obtained from TCAD simulation up to 250 GHz as shown in Fig. 7 in order to validate the parameter extraction. Table 1 summarize all the extracted parameters of the DG and SG square-shaped extended TFETs used in the verifications at $V_{GS} = 0.7$ V and $V_{DS} = 0.7$ V. As shown in Table 1 of manuscript, the values of approximations are much smaller than one even at 250 GHz ($\omega^2 R_g^2 (C_{gs}+C_{gd})^2 \approx 1.096 \times 10^{-3}$ and $3.353 \times 10^{-3}$; $\omega^2 \tau \approx 8.19 \times 10^{-3}$ and $6.44 \times 10^{-3}$ for SG and DG square-shaped extended source TFET, respectively). This verifies the validity of using the assumption in simplifying (1)-(4) to (5)-(8) in this work up to extremely high frequency range even at 250 GHz. As seen, $Y$-parameters obtained from the NQS model equivalent circuit showed close agreement with the calculation results by the TCAD simulation. The root-mean-square errors of the model were calculated to be within 3.3% and 3.7% for the DG and SG square-shaped extended TFETs, respectively. These verification results strongly support that the proposed model is accurate and valid for extended source TFETs up to the extremely high frequency range.

![Graph](image1)

**Fig. 6.** Comparison of the $f_{max}$ values between DG and SG square-shaped extended source TFET obtained from unilateral power gains as a function of $V_{OV}$.

**Table 1.** The summary of the extracted parameters ($V_{GS} = 0.7$ and $V_{DS} = 0.7$ V) for DG and SG square-shaped extended source TFET.

<table>
<thead>
<tr>
<th></th>
<th>SG square-shaped extended source TFET</th>
<th>DG square-shaped extended source TFET</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{gs}$</td>
<td>0.306fF</td>
<td>0.528fF</td>
</tr>
<tr>
<td>$C_{gd}$</td>
<td>0.513fF</td>
<td>1.233fF</td>
</tr>
<tr>
<td>$C_{sd}$</td>
<td>44.975 aF</td>
<td>14.649 aF</td>
</tr>
<tr>
<td>$R_g$</td>
<td>25.736 Ω</td>
<td>20.932 Ω</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.362 ps</td>
<td>0.321 ps</td>
</tr>
<tr>
<td>$g_m$</td>
<td>19.618 nS</td>
<td>6.817 µS</td>
</tr>
<tr>
<td>$g_{ds}$</td>
<td>34.214pS</td>
<td>52.45 nS</td>
</tr>
<tr>
<td>$\omega^2 R_g^2 (C_{gs} + C_{gd})^2$</td>
<td>$1.096 \times 10^{-3}$</td>
<td>$3.353 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\omega^2 \tau$</td>
<td>$8.19 \times 10^{-3}$</td>
<td>$6.44 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

![Graph](image2)

**Fig. 7.** Comparison of modeled (line) and values obtained from TCAD simulation (symbol) $Y$-parameters of DG and SG square-shaped extended source TFET at $V_{GS} = 0.7$ and $V_{DS} = 0.7$ V. (a) $Y_{11}$, (b) $Y_{12}$, (c) $Y_{21}$ and (d) $Y_{22}$. 
4. Conclusion

In this paper, we described RF performances of DG and SG square-shaped extended TFETs based on parameter extractions from the NQS model equivalent circuit in terms of the cut-off frequency, maximum oscillation frequency and small-signal parameters. DG square-shaped extended TFETs have higher RF performances because of the much higher transconductance than SG square-shaped extended TFETs. The results showed good agreement between the modeled Y-parameters and the extracted parameters for the high frequency range up to the cut-off frequency.

References


Saeid Marjani received the B.S. and M.S. degrees in electrical engineering from Arak Branch, Islamic Azad University, Arak, Iran, and the Ph.D. degree in electrical engineering from Ferdowsi University of Mashhad, Mashhad, Iran. His research interests include the design, modeling, fabrication, and characterization of nanoscale CMOS, tunneling FETs, optoelectronics, and photonic devices and analog integrated systems.

Seyed Ebrahim Hosseini received the B.Sc. degree in electrical engineering from the Isfahan University of Technology, Isfahan, Iran, the M.Sc. degree from Tarbiat Modares University, Tehran, Iran, and the Ph.D. degree in electrical engineering from the Sharif University of Technology, Tehran, in 2001. He is currently an Associate Professor of Electronics Engineering with the Ferdowsi University of Mashhad, Mashhad, Iran.
A Measurement Setup for Radiated EMI of Un-buffered DRAM Modules

Mojtaba Joodaki and Amir Attar

Abstract. As clock frequency and switching current of DRAM (dynamic random access memory) modules are ever increasing, electromagnetic interference (EMI) of such modules is gaining a very high degree of importance. Normally EMI measurements of memory modules are performed implementing complete motherboards or PCs. This makes it very difficult to only separate the radiated EMI of the memory module with a single EMI measurement. This is the first attempt to construct a shielded measurement setup for radiated EMI for DRAM PCBs. The measurement setup is fabricated and tested for un-buffered double data rate (DDR) modules and gives excellent results. This is very helpful in EMI and electromagnetic compatibility (EMC) investigation of memory modules. Although the measurement setup is just implemented for memory modules, the approach can be used for other types of PCBs too.

Keywords: EMC/EMI measurement, DRAM PCBs radiation.

1. Introduction

The compulsory test and evaluation of electromagnetic interference (EMI) and electromagnetic compatibility (EMC) for each electrical and electronics circuit and system include radiated emission test [1], [2]. DRAM (dynamic random access memory) modules, like any other electrical and electronic equipment and systems, are capable of emitting electromagnetic energy which can constitute electromagnetic interference (EMI). In the new memory technologies, designers are implementing effective clock frequencies as high as 3.2 GHz (DDR4-3200) and switching currents above 10 A. This generates a very high di/dt, which results in much higher radiated emissions. Furthermore, to increase the lifetime and reliability, metallic heat-sink is generally used to remove the heat from the surface of DRAM module and decrease the operating temperature of the DRAM chips on board [3], [4]. Improper implementation of such a large area of metallic component in the presence of above gigahertz signals can cause efficient antenna which degrade the EMI/EMC performance of the module drastically [5], [6]. Therefore, EMI/EMC is a major challenge for future DRAM products.

A very important point in EMI investigation of memory modules is to measure only the radiated EMI from the module itself. In this paper we shield the whole PC from outside and provide a proper slot for DRAM module by using a designed riser card. The idea is to shield the rest of the system and to measure only the fields radiated from the modules under test. This provides us with a deeper sight into the EMI/EMC performance of our memory modules. The details of the measurement setup are given in Section 3. The measurement setup is tested and the successful results are presented in Section 4.

2. Radiated EMI Regulation and Importance of EMC Consideration in Modern Electronics

A. EMC Radiated Emission Regulations

Among the radiated emissions standards available for commercial electronic products, the US standard of FCC part 15 (subpart B) and the European standard of EN 55022 (the same as CISPER 22), are the most used standards. Since digital electronic devices are prone to radiate electromagnetic waves unintentionally, every digital circuit that has a clock greater than 9 KHz, is covered by the FCC part 15 (subpart B). The two above-mentioned standards have similar test arrangements but somewhat different limits. Their limits are compared in Fig. 1.

![Radiated emission limits for FCC and CISPR measurement](image-url)

Fig. 1. Radiated emission limits for FCC and CISPR measurement distance equal to 10 m). Solid line: FCC, dashed line: CISPER [7].
The radiated emission standards cover two classes: class B for the products used in domestic environments (where the use of broadcast radio and television receivers may be expected within 10m of the apparatus concerned) while class A covers the rest of other products [8]. Since class B devices are more prone to interfere with radio and television receivers, they are more restricted (10 dB lower than class A).

B. Importance of EMC/EMI Consideration in Modern Electronics

For an electronic product to comply with an EMC standard it is needed to function properly in its intended electromagnetic environment and it should not be a source of radiation to that electromagnetic environment. Always the trend in electronic industry is toward more packed and miniaturized systems with much higher speed and more functionality but at lower cost, weight and operating voltage. Therefore the future electronic devices not only will suffer more from the EMI issues due to the small distance to the neighboring devices and much higher switching speeds, but they will be more susceptible to the external noises due to the lower operating voltage.

Table 1 shows the roadmap for the semiconductor technology development [9]. According to the roadmap as technology develops, the gate length (LG) decreases, which results in increase of the current density and the switching speed, hence, higher di/dt and higher radiated EMI. On the other hand, operating voltage (V_{DD}) reduction worsens the immunity of the devices against the EMI. This is well explained in Fig. 2.

<table>
<thead>
<tr>
<th>Year</th>
<th>Gate Length L_G (nm)</th>
<th>V_{DD} (V)</th>
<th>Intrinsic switching speed (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>18</td>
<td>1</td>
<td>2439</td>
</tr>
<tr>
<td>2011</td>
<td>16</td>
<td>1</td>
<td>2778</td>
</tr>
<tr>
<td>2012</td>
<td>14</td>
<td>0.9</td>
<td>3226</td>
</tr>
<tr>
<td>2013</td>
<td>13</td>
<td>0.9</td>
<td>3571</td>
</tr>
<tr>
<td>2014</td>
<td>11</td>
<td>0.9</td>
<td>4348</td>
</tr>
<tr>
<td>2015</td>
<td>10</td>
<td>0.8</td>
<td>4762</td>
</tr>
</tbody>
</table>

3. The Measurement Setup [10]

Far-field measurements of the radiated EMI of a complete electronic system on a PCB or motherboard is usually performed in an anechoic chamber according to the well known EMI measurement standards of FCC part 15 (subpart B) or the European standard of EN 55022 (or CISPER 22) [8]. On the other hand, far-field radiated EMI of memory integrated circuits (ICs) and packages are usually measured using transverse electromagnetic (TEM) cells under IC-EMI measurement standards such as IEC 61967-2 [11]. According to our best knowledge there is no reliable measurement method available to measure the radiated EMI of an electronic PCB or module that vertically stands on a motherboard. Therefore, this work aimed at developing such a measurement method. The required measurement setup must fulfill the following criteria:

- Though it should effectively shield the radiated EMI from the rest of the system under test, it should have minimum influence on the DRAM module emission.
- To prevent difficulties in measurements regarding transportation of the measurement setup to the EMI/EMC labs and in order to be easily handled in the semi-anechoic chamber, it should be small and light.
- Improper grounding of the motherboard and power-supply cable will results in radiation of the main shielding case which hinders proper measurements.
- It should provide a proper thermal path similar to the PC with its normal case. A shielding case with an improper thermal path will damage the system under test or influence the measured results.
- Similar to any other product, it would be in our interest to make it as cheap as possible.

A scheme of the suggested measurement setup is illustrated in Fig. 3.

In order to provide an excellent shielding and a good electrical ground for the system, the main aluminum case has a thickness of 1 mm to 3 mm. Different parts such as motherboard, power supply and ac cable are fixed and electrically connected to the main case through very low impedance paths. In order to minimize the electromagnetic interference between the power supply and the motherboard, they are separated by some aluminum partitioning walls. On the left side of the measurement setup a window is used for periphery interfaces such as monitor and keyboard cables. During the measurements, after running the memory test program, this window must be screwed up. Since the memory modules under test must be outside the main aluminum shielding case, different riser-cards and exchangeable metal cover plates are implemented for different memory module types (DDR1, DDR2, So-DIMM DDR1, and So-DIMM DDR2). There is a 3 mm wide slot in each metal cover plate for the riser-cards connector. To prevent electromagnetic emissions from these slots, 1 cm thick absorber foams and rubbers are used around the slots on the inner sides of the metal cover plates. The grilled air vents are used to transfer the dissipated heat to the outside of the setup. Openings in the air vents have a diameter of 3 mm. This provides the system with good air ventilation paths and a proper electromagnetic shielding at few GHz frequency ranges.

Fig. 2. Impact of device scaling on EMC problems.
To minimize any common mode radiation on the ac cable during the measurement, a shielded cable which is earthed to the main case is used. Fig. 4 shows the realized measurement setup prepared for radiated EMI. In Fig. 5 the top cover plate of the main case is removed to give a better view of the setup.

4. The EMI Test Results

The radiated EMI tests in this work are done according to the European standard for emission limits of EN 55022 at a certified measurement lab. The details regarding the measurement equipments and conditions are given in the reference [8]. Although this section covers only the results of unbuffered DDR modules, as mentioned above, the setup is capable of performing unbuffered So-DIMM test as well.

The clock frequency for the DDR1 and DDR2 modules were 133 MHz and 266 MHz, respectively. In order to investigate the measurement setup systematically different tests are done and presented in this section. The frequency range of interest in this evaluation is limited to 1 GHZ.

Before performing any test, to have a fair evaluation, the background EMI noise in the blank semianechoic chamber should be known. The second step is to confirm that the setup generate no considerable radiated emission from the common mode current on the surface of the main aluminium shielding case. For this purpose a metal cover plate with no slot is used to fully close the setup and during the test the system is running using an unbuffered DDR2 module inside the EMI test setup. The results of these two tests are presented in Fig. 6 and confirm no considerable self radiated emission from the test setup.
Another question raised is related to the radiated emissions of the 3 mm wide slot in the metal cover plate and the riser-card connector. At this stage, in addition to the DDR2 module inside the test setup, a metal cover plate with a slot for DDR2 modules and a DDR2 riser-card are used for the interference radiation tests. Fig. 7 shows the results and compared them with that of the fully closed test setup in Fig. 6. Regarding the results it can be easily understood that shielding at the slot which used for DDR2 riser-card is properly done.

![Graph](image1)

Fig. 6. Measured radiated emissions; (a) blank semi-anechoic chamber, (b) the fully shielded EMI test setup.

![Graph](image2)

Fig. 7. Measured radiated emissions, (a) the fully shielded EMI test setup, (b) the EMI test setup with only DDR2 riser-card placed in the slot.
Fig. 8 illustrates the interference radiation test for the unbuffered DDR2 using the new EMI measurement setup. The measured EMI is just emitted from DRAM module and the first, the second and the third harmonies of the clock frequency (266 MHz) are easily recognized. It should be noted that some EMI/EMC failed modules are intentionally prepared for this investigation to present a better understanding of the test setup. In order to present the advantages of using such a test setup and its robustness in cancelling the undesired measured emissions from the rest of the system, the same measurement with an open motherboard (with no shielding case) is performed and the result is compared with that of DDR2 with the new shielded measurement setup, see Fig. 9. The emission of the open test system at frequencies between 60 MHz to 130 MHz are mostly related to the dc power-supply system and cables and can be suppressed by using several EMI ferrite filters on the cables. In this measurement even by implementing three EMI ferrite filters still the EMI at these frequency ranges is higher than the EMI standard limit of 30 dBµV/m.

Fig. 8. Measured radiated emissions, (a) the shielded EMI test setup only with the DDR2 riser-card, (b) the shielded EMI test setup with the DDR2 riser-card and an un-buffered DDR2 PCB.

Fig. 9. Measured radiated emissions, (a) the shielded EMI test setup with DDR2 riser-card and an un-buffered DDR2 PCB, (b) the Open EMI test setup with an un-buffered DDR2 PCB.
Another important point is that the EMI information of the memory PCB under test is lost in the measurement with the open board system. This happens because of two reasons: the first is that there are stronger radiated emissions from the rest of the system and they easily cover the radiation peaks from the memory module (see the measured EMI of the DDR2 module with the open system at frequencies around 230 MHz). The second is that the radiations from the memory module are influenced by the nonlinear environment of the open system or they are easily modulated by other parts radiation and are transferred to other frequencies. This can be also seen in the measured results in Fig. 9. The EMI peak at the second harmony of the DRAM clock frequency (533 MHz) is easily distinguishable from the measured data of the DDR2 module with the shielded system; but it seems this EMI peak for the open setup is influenced by the system and transferred to a lower frequency of 500 MHz.

It is of outmost importance to prevent any improper grounding and shielding in the test setup construction. Such mistakes may cause very high radiated emissions which make proper EMI tests impossible. To have a feeling about such problems, the DDR1 memory module is measured with both open and shielded systems. In the open setup measurement no riser-card is used but in the shielded setup test a riser-card is used which is 1 cm taller than the main shielding case and is out of the case by 1 cm. Both measured results are presented in Fig. 10. The closed setup has suppressed the radiated emissions from the power-supply system and cables, however much higher unwanted emissions have been induced by imperfect shielding at the slot provided for the riser-card.

5. Conclusion

A new measurement setup for radiated emission test of un-buffered memory module was introduced and validated by systematic numbers of tests in a semi-anechoic chamber at a certified test lab. The measurements testify that the new shielded setup is very useful in EMI/EMC investigation of the memory modules and separate the radiated EMI of the memory PCB from that of the rest of the system, thus providing useful information about the EMI of the memory modules under test. Further, improper setup construction aspects are discussed and investigated experimentally. The proposed test setup can be constructed and implemented for other types of memory modules.

Acknowledgment

The author would like to thank Mr. S. Muff, Mr. U. Brandt, Mr. A. Motamedi, and Dr. M. Gospodinova from Qimonda AG for their support and useful discussions.

References


**Mojtaba Joodaki** was born in Khorram Abad, Iran in 1970. He received his B.S. and M.Sc. in electrical and electronic engineering from Iran Science and Technology University in 1994 and Tarbiat Modarres University, Tehran, Iran, in 1997 and his Ph.D. degree in electrical engineering from Kassel University, Kassel, Germany, in 2002, respectively. He joined ATMEL Germany GmbH, Heilbronn, Germany as a device engineer in 2002. There, he was working on technology development, modeling and characterization of Si and SiGe-based devices for RF applications. In April 2005, he joined Infineon Technologies AG in Munich, Germany as a development engineer, where he was responsible for EMI/EMC of memory modules. In his last industrial position, from Oct. 2006 till July 2009, he was a device engineer at Qimonda GmbH, Dresden, Germany, involved in developing nano-transistor for DRAM products. Then, he started working as a visiting scientist and lecturer at the Institute of Nanostructure Technologies and Analytics (INA) at Kassel University, where he defended his Habilitation dissertation in April 2011. Since Sept. 2010, he has been with the Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran, where he was an Assistant Professor and became an Associate Professor in Nov. 2011.

**Amir Attar** received the B.Eng. degree in robotics and automation engineering from Shahrood University of Technology in 2012 and the M.Sc. in electronics engineering from Ferdowsi University of Mashhad. He has joined the EMC lab. of Electrical Engineering Department of Ferdowsi University of Mashhad since 2013. His main area of interests is Electromagnetic compatibility, high frequency electronic, and metamaterial structures engineering.
Incorporation of Distribution System Reconfiguration and Expansion Planning Problems by Considering the Role of Demand Response Resources

Hamidreza Arasteh, Mohammad Sadegh Sepasian, and Vahid Vahidinasab

Abstract. The planning of an active distribution system is investigated in this study. This paper conducts a novel concept of smart distribution system reconfiguration and planning problem. Proposed problem uses the concept of distribution system reconfiguration (DSR) with the aim of reducing and postponing the expansion requirements, while the potential of demand response (DR) programs are considered. DR programs are modeled as virtual and distributed resources to be dealt with the distribution system expansion planning (DEP) problem in the long term time horizon. Indeed, the main purpose of this paper is to propose “demand response and distribution system reconfiguration and expansion planning (DR-DSREP)” problem to identify the impact of DSR and DR on the expansion planning of distribution systems. The 33-bus distribution system is utilized in numerical studies to investigate the performance and effectiveness of the proposed problem. The simulation results show the efficiency and advantage of the proposed methodology.

Keywords: Active distribution system, demand response, distribution expansion planning, distribution reconfiguration.

Nomenclature

Indicators:
y: planning year; 
m & n: bus number; 
m-n: line between buses m and n; 
per: time period;

Sets:
\( \mathcal{A} \): set of lines; 
\( \Delta \): set of buses; 
\( \mathcal{Y} \): set of time periods; 
\( \mathcal{Y} \): set of planning years;

Parameters:
i: discount rate;

\( U_{C_{m-n}} \): upgrading cost per kilometer of line “m-n” [$/km]; 
\( L_{m-n} \): length of line “m-n” [km]; 
\( C_{w}^{DR} \): cost of DR at bus “m” in peak period [$/kW-hour]; 
\( LC \): cost of energy losses [$/kW-hour]; 
\( t_{(per)} \): duration of each time period [hour]; 
\( V_{\text{min}}, V_{\text{max}} \): minimum and maximum permissible voltage level [kV]; 
\( I_{\text{max}_{m-n}} \): maximum current capacity of line “m-n” [A]; 
\( p_{\text{m DR (max)}} \): maximum DR capacity at bus “m” [kW]; 
\( E\text{ENS}_{\text{max}} \): maximum acceptable value of EENS [kW-hour];

Variables:
\( \text{NPVC} \): net present value of the costs [$]; 
\( C(y) \): total cost in year “y” [$]; 
\( C^U(y) \): total system upgrading cost in year “y” [$]; 
\( C^{DR}(y) \): total DR cost in year “y” [$]; 
\( C^{loss}(y) \): total cost of energy losses in year “y” [$]; 
\( n_{m-n}(y) \): number of installed lines between buses “m” and “n” in year “y”; 
\( p_{\text{max}_{m-n}}(y) \): amount of enable DR at bus “m” in year “y” [kW]; 
\( T^{DR}(y) \): DR enabling time in year “y” [hour]; 
\( p_{\text{loss}_{m-n,per}}(y) \): power losses of line “m-n” in time period “per” of year “y” [kW]; 
\( V_{m,per}(y) \): voltage level of node “m” in time period “per” of year “y” [kV]; 
\( I_{m-n,per}(y) \): current flow of feeder “m-n” in time period “per” of year “y” [A]; 
\( E\text{ENS}(y) \): value of expected energy not-supplied in year “y” [kW-hour]; 
\( p_{f_{m-n,per}}(y) \): power flow of feeder “m-n” in the time period “per” of year “y” [kW];
$z_{m-n} (\text{per, y})$: binary variable that is equal to 1 if feeder “m-n” is selected in time period “per” of year “y”; otherwise it is 0;

**Abbreviations**

- **DSR**: Distribution System Reconfiguration;
- **DR**: Demand Response;
- **DEP**: Distribution system Expansion Planning;
- **DR-DSREP**: Demand Response and Distribution System Reconfiguration and Expansion Planning;
- **GA**: Genetic Algorithm;
- **PSO**: Particle Swarm Algorithm;
- **IEA**: International Energy Agency;
- **TBP**: Time-Based Program;
- **IBP**: Incentive-Based Program;
- **MBP**: Market Based Program;
- **TOU**: Time of Use;
- **RTP**: Real-Time Pricing;
- **CPP**: Critical Peak Pricing;
- **DLC**: Direct Load Control;
- **EDRP**: Emergency Demand Response Program;
- **I/C**: Interruptible/Curtailable service;
- **CAP**: Capacity Market Program;
- **DB**: Demand Bidding;
- **A/S**: Ancillary Service;
- **LDC**: Load Duration Curve;
- **DG**: Distributed Generation;
- **EENS**: Expected Energy Not-Supplied [kW. hour].

1. Introduction

The expansion of electricity chain components is necessary due to load increases and includes generation, transmission and distribution system expansion planning. DEP is an important activity to cope with the forecasted electricity demand. A distribution network is a part of a network between distribution substations and customers’ entrance gate including distribution substations, primary distribution system, distribution transformers, and secondary distribution system [1].

DEP problem consists of sizing, timing and siting of distribution facilities, while the restrictions of the system and components are overcome [2]. This problem is necessary to satisfy forecasted load and system constraints. Distribution system planners should be able to determine the peak load amplitude and its location to provide a suitable and efficient expansion plan with the optimal cost [3]. DEP methods have been assessed through numerous studies and various optimization algorithms are proposed to solve the introduced problems [4]-[6]. Network expansion planning requires a complex optimization procedure due to the nonlinear and combinatorial nature of the problem [7], [8]. Therefore, various studies are focused on utilizing some methods with random nature. However, such algorithms cannot guarantee the global optimum solution that is the main drawback of these methods [9], [10].

As mentioned above, lots of studies have investigated single or multi-stage expansion problem of distribution systems with the aim of minimizing the investment and operation costs [11], [12]. Optimization algorithm should be employed for the best allocation of the limited financial resources [11]. Pseudo-dynamic theory [13], dynamic planning [14], graph-theory models [15] and heuristic algorithms such as GA are examples of the introduced optimization methods. According to the trends of studies in recent years, heuristic methods are being used increasingly in spite of their random nature [16], [17].

Recent studies and reports strongly focused on the importance and necessity of smart grids [18]. DR is considered as the core of smart grids and enabled by end-users to motivate changes in power consumption patterns. Reference [19] has investigated the influence of DR programs in a future smart electricity system in 2020. The effectiveness of DR programs has been assessed in [20]. According to the strategic plan of IEA, DR programs are considered as the first choice in all energy policy decisions [21]. The potential benefits of the demand side activities are introduced as a reason of such considerations [21]. Recently, DR programs attracted more attentions and are considered as resources, called DR resources. DR programs can be divided into three major classes, including [16]: TBP, IBPs, and MBPs. TBP programs consist of TOU, RTP, and CPP. In these programs, customers should cope with varying levels of time-dependent prices; the least with TOU and the most with RTP. IBPs consist of DLC, EDRP, I/C, and CAP. EDRPs are voluntary programs in which customers are not penalized if they do not respond to the DR calls. In the DLC programs, utilities can directly curtail customers’ electricity using a remote switch. I/C and CAP are mandatory programs and they use penalties if enrolled customers do not reduce their consumptions when directed. MBPs include DB and A/S programs. In DB programs, large customers will be encouraged to have participation at their desired price, or to determine that how much load they are willing to curtail at a specific posted price. In A/S programs, customers are allowed to bid load curtailment in electricity markets as operating reserves [16].

As it is explained in [22], distribution companies are one of the most important buyers for DR resources. They can purchase DR resources in a regulated market-based or conventional bilateral manner. From the economic point of view, distribution planners want to minimize all the investment costs (long-term horizon) as well as the operational costs (short-term horizon). Although DR programs are substantially short term activities, their effects on yearly LDC is not negligible. Indeed, in addition to the changes of daily electricity pattern, they are able to modify the LDC. Consequently, DR programs can be motivated in short-term time horizon; while the long-term aims are considered. In [23], the authors have investigated the effects of the DR programs on the planning of distribution systems.

DSR is another component of active distribution systems that is investigated in this paper. Distribution systems have some normally close and normally open switches. By changing the state of the switches, the configuration of the system will be changed. Generally, reconfiguration is to transfer parts of loads from one feeder to another. Lines’ power flow, power losses, and voltage levels change via switching operations. So, DSR can reduce power losses and improve the operational condition of the system.
Furthermore, releasing the capacity of the transmission and distribution networks as well as the substations capacity can reduce the expansion requirements as a consequence of DSR. Thereby, DSR can have a direct effect on the expansion plans. It should be mentioned that by using the DSR, the state of the switches is changed. Consequently, in some time intervals, some of the corridors are not operated, while in some other time intervals, they may be used. Hence, by the reconfiguration of the system, some corridors are not operated just in some specific time intervals, but they still exist in the system. Indeed, distribution systems could be designed as meshed networks but they should be operated radially by opening some of the switches.

A lot of studies have investigated the problem of DSR [24], [25]. A long-term multi-objective planning framework is proposed in [26] to maximize the benefits of DSR beside the allocation of distributed generation units. Lines’ reinforcement plan, network reconfiguration and the planning of distributed generation units are handled in [26]. A heuristic reconfiguration algorithm is presented in [27] that minimize the non-delivered power in contingency conditions.

A wide range of algorithms has been introduced to solve the DSR problem. A method based on bacterial foraging optimization algorithm is presented in [28] with the aim of loss minimization. Moreover, improved adaptive imperialist competitive algorithm [29], GA [30], artificial immune systems [31], and some other methods such as classical optimization techniques, parallel simulated annealing, reactive tabu search, database, and knowledge-based heuristic algorithms are proposed until now [27].

Basically, the DEP is a problem that considers the expansion and operation cost terms, besides the reliability of the system. In fact, the utility should provide a cost-effective and reliable service to provide the electricity demand with a standard quality level [32] Reference [33] presented a multiobjective problem for the DEP by considering the planning costs and a reliability index (energy not served).

An MINLP model is proposed in [34] for the DEP problem, considering the expansion and operation cost terms as well as the reliability costs. In this paper, the reliability costs are computed by calculating the non-supplied energy in the distribution network. El-Zonkoly et al. [35] utilized the comprehensive learning PSO (CLPSO) to minimize the generation cost as well as un-served power cost. Zou et al. [36] introduced an analytical method to access the desired reliability of distribution systems based on the following criteria: system average interruption duration index (SAIDI) and system average interruption frequency index (SAIFI). The presence of DGs (dispatchable and nondispatchable renewable DGs) is also considered in this paper. Reference [37] proposed a time-sequential simulation approach to compute the cost of reliability in distribution systems. Chowdhury et al. [38] investigated the reliability level of distribution systems by considering the presence of conventional DGs like gas turbines. Several multiobjective problems have been presented to model the multistage DEP under dynamic or pseudodynamic procedures [39]. Consequently, optimization methods like MOPSO, NSGA-II, and NBI have been introduced to solve multiobjective problems [40]-[42]. MOPSO is one of the appropriate meta-heuristic methods in order to solve the complex optimization problems due to its robustness in controlling parameters and its flexible applications [43].

Considering the importance of this research area, this paper proposes the “DR-DSREP” problem which incorporates DR, DSR, and the expansion planning problems.

As it is known, the presence of DGs is one of the most important components of the smart grid. However, the aim of this paper is to investigate the effect of DR and DSR in the planning of distribution systems. As it is mentioned, DR and DSR are essential components to construct an active distribution system. Hence, this paper investigates the necessity of DR, DSR, and the integration of them, in the DEP problem. The presence of DG units and their integration in the system can be investigated through another comprehensive study.

Regarding the role of DR resources in the future smart grid, DR models are developed in this paper to be dealt with the long-term planning framework. Hence, the effect of DR programs on the LDC is modeled and investigated through this paper. It should be mentioned that, for the sake of simplicity and without loss of generality, the DLC programs are considered in this paper to avoid the probabilistic nature of DR. However, stochastic models can be developed to evaluate the effect of other types of DR which is beyond the scope of this paper.

As aforesaid, DSR is addressed in the current study as another part of active distribution systems with the aim of reducing and postponing the expansion investments. Therefore, the decision variables of the reconfiguration problem are considered besides the DEP and DR variables. In order to satisfy the system reliability level, the EENS index is considered in this paper. A pseudo-dynamic procedure is utilized to solve the multi-stage problem. Furthermore, PSO is assigned in this paper to optimize the proposed problem.

In regard to previous studies, the main difference between this paper and [23] is to integrate the simultaneous effects of DSR and DR, on the DEP problem. Hence, by using some suitable scenarios, the effectiveness of DR and DSR, and also the integrated effects of them on the expansion planning of distribution systems are investigated. Furthermore, the EENS index is considered in this paper as the reliability criteria of the system as one of the system constraints.

The main contributions of the paper can briefly be classified as:
- Providing the long-term model of DR as some virtual distributed resources;
- Considering the role of DR resources in the DEP problem;
- Incorporating DSR with DEP to minimize total upgrading cost, while the potential of DR programs is modeled.

The rest of the paper is organized as follows. Problem description is explained in Section 2 which consists of the mathematical formulation of the objective function, system constraints, power flow, and optimization algorithm. Section 3 conducts the simulation results. Finally, concluding remarks are drawn in Section 4.
2. Problem Description

In this section, the problem is mathematically formulated. The multi-stage planning problem is considered in the long-term time horizon and solved by using the PSO method. The details of the PSO algorithm are explained in Section 2.4. Integrating the potential of DR and DSR in the DEP is the specific feature of the proposed problem.

As explained in section 1, DR programs can change the shape of LDC. Fig. 1 schematically illustrates the effect of DR resources on the LDC. As shown in Fig. 1, after enabling DR in a network, a part of peak demand will be transferred to the shoulder and off-peak periods. Loads can be classified as multi-period or single-period loads.

In multi-period loads, a specified percentage of loads can transfer to other periods. Single-period loads cannot be shifted to other times and they should be turned off when they are called to participate in DR programs. The modified LDC after considering DR in the network is depicted using dashed lines in Fig. 1.

Mathematical formulation of the problem is given as follows.

A. Objective Function

The objective function is to minimize the total expansion costs and simultaneously satisfy system constraints and cope with the forecasted load. The mathematical formulation can be expressed as equation (1).

\[
\min \left\{ NPVC = \sum \frac{1}{(1+i)^{y}} \times C(y) \right\}
\]

\[
C(y) = C^U(y) + C^{DR}(y) + C^{Loss}(y)
\]

\[
C^U(y) = \sum_{m-n} \left( UC_{m-n} \times L_{m-n} \times n_{m-n}(y) \right)
\]

\[
C^{DR}(y) = \sum_{m} \left[ C^{DR}_{m} \times p^{DR}_{m}(y) \times T^{DR}(y) \right]
\]

\[
C^{Loss}(y) = \sum_{per} \sum_{m-n} \left[ p^{Loss}_{m-n,per}(y) \times t(\text{per}) \times LC \times z_{m-n}(\text{per}, y) \right]
\]

\[
\forall \ m-n \in \Lambda, \ m \& n \in \Delta, \ \text{per} \in \Upsilon, \ y \in Y
\]

In (1), \( C(y) \) denotes the required upgrading cost of the feeders. \( C^{DR}(y) \) is the total cost that is needed to enable DR in the network. Moreover, \( C^{Loss}(y) \) is the total cost related to the energy loss in the network.

Decision variables in the objective function formulation are line reinforcement \( (n_{m-n}(y)) \), system reconfiguration \( (z_{m-n}(\text{per}, y)) \), and the specification of DR programs \( (p^{DR}_{m}(y)) \). Furthermore, NPVC, \( C(y) \), \( C^U(y) \), \( C^{DR}(y) \), \( C^{Loss}(y) \), \( T^{DR}(y) \), and \( p^{Loss}_{m-n,per}(y) \) are other variables of the objective function.

B. Constraints

Problem restrictions can be classified as follows.

1) Radiality and Connectivity of the Network

Distribution systems have tree shape graphs and should be operated radially [44]. Hence, islanded buses should not appear for providing the system loads. So, all the nodes in a fully connected tree shape distribution networks should be connected to the root of the graph [45], [1]. Furthermore, if system graph is connected, not islanded, equation (2) should be satisfied to ensure the radiality of the system. In (2), \( N^l \) and \( N^n \) are the number of network lines and buses, respectively. The presented approach in [46] is utilized in this paper to guarantee the radiality and connectivity of the network.

\[
N^l = N^n - 1 \quad (2)
\]

2) Permissible Voltage Levels

Voltage levels should not exceed the acceptable ranges. So, voltage constraints should be applied as inequalities (3).

\[
V_{min} \leq V_{m-per}(y) \leq V_{max}, \ \forall \ m \in \Delta, \ \text{per} \in \Upsilon, \ y \in Y \quad (3)
\]

3) Current Limits

The maximum current limits of lines are represented by (4).

\[
-I_{max} \leq I_{m-n,per}(y) \leq I_{max}, \forall m-n \in \Lambda, \ per \in \Upsilon, \ y \in Y \quad (4)
\]

4) The Maximum Penetration Level of DR

DR programs have limitations because of their barriers including customer barriers, producer barriers, and structural barriers [47]. These barriers are discussed in [47] by details. However, DR penetration level constraints are simply formulated by inequalities (5).

\[
I^{DR}_{m}(y) \leq p^{DR}_{m}(y \text{max}), \forall m \in \Delta, \ y \in Y \quad (5)
\]

5) Reliability Constraint

The reliability of the network should maintain in the acceptable range. The EENS index is utilized in this paper as the reliability criterion, as formulated by (6) and (7).
\[ EENS(y) \leq EENS^{\text{max}}, \quad \forall \quad y \in \Psi \]  

(6)

in which

\[ EENS(y) = \sum_{p \in \pi} \sum_{y \in \Psi} \left[ \lambda_{m,n} \times r_{p \text{, } m,n} \times \left( \frac{t_{\text{per}}}{8760} \times L_{m,n} \right) \times p_{f_{m,n}} \times (p_{\text{per}}, y) \times z_{m,n} (p_{\text{per}}, y) \right] \]  

(7)

where \( \lambda_{m,n} \) and \( r_{p \text{, } m,n} \) are the failure rate of feeder “m-n” per kilometer and per year (in fail/km. year) and the average duration of fault on feeder “m-n” (in hr/fail), respectively.

C. Power Flow

The backward/forward sweep method is utilized in this paper for power flow calculation. The algorithm of this method is shown in Table 1.

Fig. 2 shows a typical radial distribution system with \( N \) load points. \( Z_m \) indicates the impedance of line between nodes “m-1” and “m”. \( I_{m,l,m} \) and \( L_{m} \) are the currents of the main and lateral lines emanated from node “m” respectively. The substation voltage level is denoted by \( U_0 \).

\( P_n \) and \( Q_n \) are the active and reactive load levels of each load points. The procedure of the introduced method can mathematically be formulated with (8)-(14). Index \( \nu \) denotes the iteration number of the backward/forward sweep algorithm (\( \nu = 1, 2, ... \)).

1) Initializing Step

\[ \nu = 1, U_n^{(\nu-1)} = U_0, \quad \forall \quad n = 1 : N \]  

(8)

2) Backward Process

Table 2 shows the formulation of a backward process. \( P_{n}^{\text{DR}} \) and \( Q_{n}^{\text{DR}} \) in Table 2 are enabled active and reactive powers with DR programs at bus “n”. \( P_{n}^{\text{b}} \) and \( Q_{n}^{\text{b}} \) are the active and reactive powers that are shifted from other periods to this period as a result of DR. Furthermore, \( U_n^{*(\nu-1)} \) is the conjugate of \( U_n^{(\nu-1)} \).

3) Forward Process

The forward process can be formulated as (14) to compute the voltage of each bus in iteration “\( \nu \)”.

\[ U_n^{(\nu)} = U_n^{(\nu-1)} - Z_n \times I_n^{(\nu)}, \quad \forall \quad n = 1 : N \]  

(14)

<table>
<thead>
<tr>
<th>Table 1. Backward/forward sweep algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

Fig. 2. A typical radial distribution system.

Table 2. The backward-sweep formulation

<table>
<thead>
<tr>
<th>Equation Numbers</th>
<th>The Absence of DR</th>
<th>The Presence of DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(9)</td>
<td>( I_{L,n}^{(\nu)} = \overline{S}_n \times U_n^{(\nu-1)}, \quad \forall \quad n = 1 : N )</td>
<td>( I_{L,n}^{(\nu)} = \overline{H}_n \times U_n^{(\nu-1)}, \quad \forall \quad n = 1 : N )</td>
</tr>
<tr>
<td>(10)</td>
<td>( I_{M,n}^{(\nu)} = \sum_{b=1}^{N} I_{L,b}^{(\nu)}, \quad \forall \quad n = 1 : N )</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>( \overline{S}_n = P_n - jQ_n, \quad \forall \quad n = 1 : N )</td>
<td></td>
</tr>
<tr>
<td>(12)</td>
<td>( \overline{H}<em>n = \left( P_n - P</em>{n}^{\text{DR}} + P_{n}^{\text{b}} \right) - j \left( Q_n - Q_{n}^{\text{DR}} + Q_{n}^{\text{b}} \right), \quad \forall \quad n = 1 : N )</td>
<td></td>
</tr>
<tr>
<td>(13)</td>
<td>( \overline{Z}_n^{(\nu-1)} = \frac{1}{U_n^{*(\nu-1)}}, \quad \forall \quad n = 1 : N )</td>
<td></td>
</tr>
</tbody>
</table>
The process should be repeated by substituting $v_{j+1} = v_{j} + 1$ until the satisfaction of the convergence criteria [23].

3. Solution Method

The PSO technique is implemented to solve the optimization problem. The PSO is a population-based optimization algorithm introduced by Eberhart and Kennedy [48]. It is based on the number of particles and inspired by the behavior of insects’ swarm or birds’ flock [49]. The PSO has some important advantages in comparison with other heuristic methods like GA. The PSO has more effective memory capacity, more diversity to search the optimum solution and also faster search speed [50].

Swarms in the PSO algorithm consist of the group of particles that determine the solution points [50]. Each particle moves in the solution space toward the best solution with a specific velocity, while it has memory to save its best previous position [48]. The $i^{th}$ particle velocity is assigned based on (15).

$$\zeta_{i}(j+1) = \zeta_{i}(j) + r_1 \times (G(j) - x_{i}(j)) + r_2 \times (P(j) - x_{i}(j))$$ (15)

where, “$j$” represents the number of iterations, and $\zeta_{i}$ expresses the velocity of particle “$i$”. “$r_1$” and “$r_2$” are random variables between 0 and 1. $G(j)$ is the best solution of all particles (global best solution) until the iteration number “$j$”. $P(j)$ is the best solution of the $i^{th}$ particle (individual best position) until the iteration “$j$”. Furthermore, $x_{i}$ denotes the position of the $i^{th}$ particle in the solution space. According to (15), the velocity of each particle in the PSO method is based on its current and previous conditions and also the positions of other particles. The decision variables can be updated based on (16).

$$x_{i}(j+1) = x_{i}(j) + \zeta_{i}(j+1)$$ (16)

Line reinforcement, network reconfiguration plan, and the specification of DR programs are considered as the decision variable of the proposed problem that should be determined using the PSO algorithm. Table 3 illustrates all the equations regard to the decision variables of the proposed problem. Equations [17]-[22] in this table formulate the velocity and position of the swarms of these decision variables. Fig. 3 shows the algorithm of the proposed problem. In this figure, $P^o (p,j)$ is the best

<table>
<thead>
<tr>
<th>Decision Variables</th>
<th>Equations</th>
<th>Equation Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line reinforcement</td>
<td>$DV_{i}^{r}(j+1) = DV_{i}^{r}(j) + v_{i}^{r}(j+1)$</td>
<td>(17)</td>
</tr>
<tr>
<td></td>
<td>$v_{i}^{r}(j+1) = v_{i}^{r}(j) + r_1 \times (G^{r}(j) - DV_{i}^{r}(j)) + r_2 \times (P^{r}(j) - DV_{i}^{r}(j))$</td>
<td>(18)</td>
</tr>
<tr>
<td>Demand response</td>
<td>$DV_{i}^{dr}(j+1) = DV_{i}^{dr}(j) + v_{i}^{dr}(j+1)$</td>
<td>(19)</td>
</tr>
<tr>
<td></td>
<td>$v_{i}^{dr}(j+1) = v_{i}^{dr}(j) + r_3 \times (G^{dr}(j) - DV_{i}^{dr}(j)) + r_4 \times (P^{dr}(j) - DV_{i}^{dr}(j))$</td>
<td>(20)</td>
</tr>
<tr>
<td>Network reconfiguration</td>
<td>$z_{i}(j+1) = z_{i}(j) + v_{i}^{z}(j+1)$</td>
<td>(21)</td>
</tr>
<tr>
<td></td>
<td>$v_{i}^{z}(j+1) = v_{i}^{z}(j) + r_5 \times (G^{z}(j) - DV_{i}^{z}(j)) + r_6 \times (P^{z}(j) - DV_{i}^{z}(j))$</td>
<td>(22)</td>
</tr>
</tbody>
</table>
solution of swarm “p” until the “jth” iteration. Moreover, \( G^p_j \) denotes the best solution of all the swarms until the iteration number “j”. \( I_{\text{max}} \) is the maximum number of the PSO iterations. If each generated particle cannot satisfy the system restrictions, a penalty factor will be considered for it. Consequently, undesirable solutions will be avoided because of the high amount of penalties.

Next section provides all the simulation results and comprehensive analysis that is required to investigate the performance of the proposed problem.

4. Numerical Results

A. Input Data and Assumptions

The 33-bus distribution system is considered as a case study to explore the planning results. Fig. 4 illustrates the primary configuration of this system. Dashed lines in Fig. 4 are related to the tie-lines and colored circles determine the candidate buses that have a potential to participate in DR programs. System specifications are provided in Table 4. All the 33-bus test system data are adopted from [51]. The maximum capacity of lines is assumed to be 118 A. The multi-stage problem is solved using the pseudo-dynamic approach. The base standard load data is considered as the forecasted demand at stage 1 (the first year of planning horizon). It is supposed that the load levels are increased by 10 percent per year with respect to the previous year for buses 8-18, and 5 percent for remained buses. It should be mentioned that the salvage value of the lines are related to their lifetimes that are usually more than the horizon planning time of the distribution systems that is considered in this paper. Hence, by considering the main aim of this paper that is to show the effects of the DR and DSR on the DEP problem, and without loss of generality, the wires lifetime is not considered in this paper. Furthermore, according to [26], the switching cost is around 203 ($/switching). It is clear that the switching cost is negligible in comparison with other cost terms in the distribution expansion problem. Hence, the switching costs can be neglected without affecting the validity and accuracy of the results.

Both primary and modified LDCs after considering the effect of DR programs are computed for each node of the system as it is depicted in Fig. 1. In Fig. 1, horizontal dashed lines are corresponding to the modified LDC. Vertical arrows denote the changes of load levels in each time segment after DR implementation. Three load levels are considered as the primary LDC segments including peak, shoulder and off-peak periods. The durations of segments are supposed equal to 360, 4900 and 3500 hours, respectively. In addition, demand levels in shoulder and off-peak periods are considered to be 75 and 60 percent of the peak load data. Also, DLC programs are considered here to avoid the probabilistic nature of DR programs. According to Fig. 1, by considering the effect of DR resources, estimated 3-segment LDC will break into 5-segment curve. Intervals [T1-T2] and [T3-T4] are equal to [0-T1] which is total hours that DR programs are enabled in the network.

As it is mentioned in the previous section, multi-period loads can be shifted to other periods, while single-period loads cannot. If DR is enabled in peak periods, the amount of load in peak time will be decreased, while the load level in other periods will be increased due to the shifted loads. So, some parts of curtailed load will be transferred to shoulder, some parts will be shifted to off-peak, and some parts will be turned off. It is assumed that 20 percent of the curtailed load will be transferred to shoulder, some parts will be shifted to off-peak, and some parts will be turned off. It is assumed that 20 percent of the curtailed load will be transferred to the shoulder hours and 50 percent of responsive load will be shifted to off-peak area. Residual loads are considered to be turned off (single period loads). Furthermore, the value of discount rate is considered equal to 20% to compute the net present value of the cost terms.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Dimension</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage base</td>
<td>kV</td>
<td>12.66</td>
</tr>
<tr>
<td>Energy cost</td>
<td>$/MW.hour</td>
<td>60</td>
</tr>
<tr>
<td>Cost of DR</td>
<td>$/MW.hour</td>
<td>200</td>
</tr>
<tr>
<td>Lines' reinforcement cost</td>
<td>$/km</td>
<td>145000</td>
</tr>
<tr>
<td>planning horizon time</td>
<td>Year</td>
<td>4</td>
</tr>
<tr>
<td>Substation voltage</td>
<td>Per-unit</td>
<td>1.04</td>
</tr>
<tr>
<td>permissible voltage levels</td>
<td>Per-unit</td>
<td>[0.96, 1.04]</td>
</tr>
</tbody>
</table>

Fig. 4. The 33-bus distribution system.
B. Simulation Results and Analysis

As it is listed in Table 5, four scenarios are defined here to explore the effectiveness and performance of the proposed problem, including:

- **Scenario 1**: Distribution system expansion planning;
- **Scenario 2**: Distribution system expansion planning considering the potential of DR resources;
- **Scenario 3**: Integrated distribution system reconfiguration and expansion planning;
- **Scenario 4**: Demand response and distribution system reconfiguration and expansion planning.

The first scenario optimizes the expansion planning problem in a conventional manner without considering the potential of DR resources and DSR. The second scenario takes into account the potential of DR resources in the planning problem. Scenario 3 integrates the problem of DSR with expansion planning of the distribution system. Finally, Scenario 4 is elaborated to investigate the effectiveness of the proposed DR-DSREP problem in which DEP and DSR problems are incorporated in the presence of DR resources. Table 6 describes the cost details of simulation results for all the scenarios. As it can be seen in Table 6, the net present value of the base case (scenario 1) is 1.85 (M$). It is decreased to 1.63, 1.48 and 1.00 (M$) and shows 11.89, 20.00 and 45.95 percent cost reduction using scenarios 2, 3 and 4, respectively. Open sections in each year are listed in Table 7. It should be mentioned that

### Table 5. Different scenarios

<table>
<thead>
<tr>
<th>Scenario Numbers</th>
<th>DR Programs</th>
<th>DSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1 (#1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario 2 (#2)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Scenario 3 (#3)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Scenario 4 (#4)</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Table 6. Details of the planning cost for each scenario

<table>
<thead>
<tr>
<th>Scenario Numbers</th>
<th>Objective Function (M$)</th>
<th>Cost Terms (M$)</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.85</td>
<td>System upgrading cost</td>
<td>1.02</td>
<td>0.07</td>
<td>0.95</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total cost of energy loss</td>
<td>0.029</td>
<td>0.033</td>
<td>0.037</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cost of DR programs</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>1.63</td>
<td>System upgrading cost</td>
<td>0.87</td>
<td>0.0</td>
<td>0.36</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total cost of energy loss</td>
<td>0.027</td>
<td>0.031</td>
<td>0.038</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cost of DR programs</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.007</td>
</tr>
<tr>
<td>3</td>
<td>1.48</td>
<td>System upgrading cost</td>
<td>0.58</td>
<td>0.36</td>
<td>0.36</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total cost of energy loss</td>
<td>0.031</td>
<td>0.028</td>
<td>0.035</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cost of DR programs</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>System upgrading cost</td>
<td>0.43</td>
<td>0.00</td>
<td>0.51</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total cost of energy loss</td>
<td>0.028</td>
<td>0.029</td>
<td>0.034</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cost of DR programs</td>
<td>0.002</td>
<td>0.009</td>
<td>0.002</td>
<td>0.016</td>
</tr>
</tbody>
</table>

### Table 7. Open sections and new added lines in each scenario

<table>
<thead>
<tr>
<th>Scenario Numbers</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Added Lines 1-2 (×2), 2-3 (×2), 3-4, 4-5, 5-6, 6-7</td>
<td>1-2</td>
<td>1-2, 3-4, 4-5, 5-6, 6-7</td>
<td>1-2, 3-4, 4-5, 5-6, 6-7</td>
</tr>
<tr>
<td></td>
<td>Added Lines 1-2 (×2), 2-3 (×2), 3-4, 4-5, 5-6</td>
<td>-</td>
<td>1-2, 3-4, 4-5</td>
<td>1-2, 3-4, 4-5</td>
</tr>
<tr>
<td></td>
<td>Added Lines 1-2 (×2), 2-3, 3-4, 4-5</td>
<td>1-2, 2-3, 3-4, 4-5</td>
<td>6-26</td>
<td>1-2, 2-3, 3-4, 4-5, 5-6</td>
</tr>
<tr>
<td></td>
<td>Added Lines 1-2 (×2), 2-3, 3-4</td>
<td>-</td>
<td>1-2, 2-3, 4-5, 5-6</td>
<td>6-26, 26-27</td>
</tr>
</tbody>
</table>
Table 7 represents the corridors with open switches in each year. It means, for instance, in the first year of the scenario 4, by using switching operations, corridors 7-8, 10-11, 14-15, 32-33, and 25-29 will be opened and so will not be operated, while all other corridors are operated by closing the corresponding switches. Moreover, it should bear in mind that, as it is mentioned, the first scenario optimizes the DEP problem without considering DR and DSR. Hence, there is not any change in the states of the switches in this scenario. So, as it is obvious in Table 7, the opened switches are always same (because the potential of DSR is not considered in this scenario). However, to satisfy the increasingly load level of distribution buses, distribution feeders are upgraded. Hence, in this case, there is not any change in the state of switches because DSR is not applied and switches of tie-lines are always open, but distribution lines are upgraded to satisfy the system constraints during the planning years. Fig. 5 shows the system reconfiguration and reinforcement plan in each year of the planning years for the scenario 4. Dashed lines in this figure indicate the upgraded corridors and the number of added lines.

The comparison results are shown in Fig. 6 for all the introduced scenarios. It can be concluded from Fig. 6, considering the potential of DR resources in the planning studies and incorporating DSR with DEP can reduce and postpone a major part of expansion costs and provide economic benefits for distribution system planners. According to Fig. 6, the proposed problem does not show any high investment requirements during all the planning studies. So, reducing a major part of expansion costs by DR

![Fig. 5. The reconfiguration and expansion scheme of the system for the scenario 4.](image)

![Fig. 6. Present value of total planning cost in each stage of each scenario.](image)
and DSR and postponing some parts of costs to the later years, decreases the net present value of total costs. The amounts of cost terms are shown individually in Fig. 7 for each scenario. Furthermore, Fig. 8 illustrates the share of each term as a percentage of the total expansion cost.

According to Fig. 7, integrating DR and DSR with the expansion planning problem will dramatically reduce the system upgrading costs. As the main aim of the proposed problem is to reduce the total expansion costs, incorporating DR, DSR and DEP have higher effect on the required upgrading costs. Thereby, comparing Figs. 8-a and 8-d expose that the upgrading cost of lines has less percentage in the proposed problem than the first one. This fact is also correct for other scenarios that separately show the effect of DR and DSR. However, the penetration level of DR is impressive in the planning results. It should be noted that, in this paper, the DR potential is considered to be very limited, i.e. DR is just applicable on buses 13-18 and 30-33 and just 20 percent of the load in each bus can participate in DR programs. The effect of the other penetration levels of DR on the planning cost is depicted in Fig. 9 for the proposed problem.

![Fig. 9. Total planning cost with respect to the DR penetration level.](image)

![Fig. 7. Present value of total lines' upgrading cost, total loss cost, and total cost of DR in different scenarios.](image)

![Fig. 8. The share of cost terms in the total planning costs: a) scenario 1, b) scenario 2, c) scenario 3, d) scenario 4.](image)
Fig. 9 illustrates the changes of the expansion cost with respect to the various penetration level of DR in the network. Furthermore, Fig. 10 represents the voltage levels of the 33-bus distribution system for each scenario in peak periods. It should be noted that the lower permissible voltage level is considered equal to 0.96 per-unit in this paper, which is completely satisfied in all the planning years. Finally, according to all the numerical analysis, the simulation results approve the efficiency and advantages of the proposed problem.

5. Conclusion

The active distribution system expansion planning is addressed in this paper by proposing “demand response and distribution system reconfiguration and expansion planning (DR-DSREP)” problem. DR programs are modeled as some virtual and distributed resources that can be used in a long-term expansion planning. Furthermore, DSR is integrated with DEP problem with the aim of reducing and postponing the total expansion cost. The concurrent impacts of DR and DSR are fully investigated as the main purpose of this study. The proposed problem is tested using the 33-bus distribution system and analyzed through four individual scenarios. The simulation results approved the economic benefits of the proposed problem as well as the operational advantages.

References


Hamid Reza Arasteh was born in Zanjan, Iran, in 1988. He received the B.Sc. degree in power engineering from the Tabriz University, Tabriz, Iran, in 2010, the M.Sc. degree from Tarbiat Modares University (TMU), Tehran, Iran, in 2012. He is currently pursuing the Ph.D. degree in power engineering at Shahid Beheshti University (SBU), Tehran, Iran. His research interests include distribution system planning, demand response, smart grids, and electricity market.

Vahid Vahidinasab received the B.Sc. degree from the K. N. Toosi University of Technology, Tehran, Iran, in 2004, and the M.Sc. (Hons.) and Ph.D. degrees from the Iran University of Science and Technology, Tehran, in 2006 and 2010, respectively, all in electrical engineering. He is currently an Assistant Professor with the Department of Electrical Engineering, Shahid Beheshti University, Tehran. His current research interests include operations and economics of electric energy systems and smart microgrids.
Mohammad Sadegh Sepasian was born in Tehran, Iran, in 1967. He received the B.Sc. degree from Tabriz University, Tabriz, in 1990 and the M.Sc. and Ph.D. degrees from Tehran University and Tarbiat Modares University, Tehran, in 1993 and 1999, respectively. Currently, he is an Assistant Professor with the Department of Electrical Engineering, Shahid Beheshti University, Tehran. His research interests include power system planning as well as distribution system planning issues.
Study on Health Monitoring of Concrete Structures Using Wireless Sensor Networks

Saman Shoorabi Sani, Majid Baghaei-Nejad, and Mona Kalate Arabi

Abstract. In this study, a system for monitoring the structural health of bridge deck and predicting various possible damages to this section was designed based on measuring the temperature and humidity with the use of wireless sensor networks then it was implemented and investigated.

This paper also presents the experimental development of an automatic wireless sensor monitoring system for concrete structures. The objective is to provide a solution to measure both temperature and humidity inside a concrete structure. The research has been focused in the early age and curing phase period. Four solutions have been addressed.

The first one involves the use of a negative temperature coefficient thermistor and an IRIS mote allowing for the creation of an IEEE 802.15.4 network. The second one considers the use of the SHT15 sensor, together with the PIC18F4680 microcontroller or the Arduino platform. The third solution involves the use of the SHT21S sensor and the ez430-RF2500 wireless development tool platform for the MSP430 microcontroller. Finally, the fourth solution considers both the SHT15 and SHT21S sensors completely shielded allowing for the creation of a long-term solution. The potential of applying the proposed inexpensive wireless sensor network approach is completely investigated and verified.

Keywords: Structural health monitoring, wireless sensor networks, concrete, sensor, temperature, humidity.

1. Introduction

Maintaining the safety and reliable service of a large bridge over its relatively long life requires obtaining continuous and reliable data regarding its structure, including the damages which are caused by the temperature gradient, cracking, fatigue, corrosion of structures and the decrease of load capacity of the bridge, etc. They all should be carefully evaluated. Common measurements such as periodic visual inspections and controlled loading test are typical in this respect and their limitations and disadvantages have been thoroughly investigated.

A new technology called structural health monitoring (SHM) which use wireless sensor networks [1]-[8] has recently attracted a lot of attention in the field of measurement and analysis of those mentioned factors. There are various SHM systems that can detect damages of the bridge structure through analyzing the dynamic characteristics of the bridge [9], [10].

In this study, a SHM system which utilizes the wireless sensor networks (WSN) and is based on monitoring humidity and thermal responses of environment has been designed and analyzed with the help of a hypothetical bridge. Researchers claim that this method has the ability to bypass the problems and limitations of the other SHM methods [11], [12]. Exposure to sun and heat exchange with the environment leads to temperature differences in different parts of the bridge. Such changes occur continuously and slowly every day and affect the structure of the bridge [13], [14]. The temperature difference between the different parts leads to the thermal response of the bridge including thermally induced strains, stresses, and changes in the reactions of bridge piers [15]. The change of these responses is slow, so they can be easily distinguished from the thermal responses caused by temporary traffic. Furthermore, they have many measureable effects. In the case of pre-stressed concrete bridges, thermally induced stresses are usually in the same range of live load stresses but are often greater than these stresses [16]-[22].

In view of the above discussion, a SHM based on environmental thermal responses seems suitable for long bridges with several curves. The main focus of this research is on pre-stressed reinforced concrete bridges with medium to high curve lengths. This study also examines the evaluation and monitoring of effects of humidity on different parts of the bridge and also the corrosion and damage caused by humidity or those damages for which humidity act as accelerating factor. The results showed that a well-designed and well-implemented SHM system based on environmental thermal responses and humidity has the ability of detecting structural damages and identifying their location and their severity [23]-[25].

The second section of this article introduces the details of proposed SHM system and its implementation, hardware, configuration of sensor network and developed monitoring program. The third section, which is the primary objective of this interdisciplinary research, is to develop a prototype for Wireless Sensor Networks allowing for remotely monitoring certain concrete structures , through testing several kinds of sensors that have the capability to be matched with the proposed SHM system and investigating their properties, accuracy, reliability and limitations in SHM of concrete structures.

From the application perspective, WSNs are useful in situations that require quick or infrastructure-less
Temperature is an important parameter during the curing and hardening of the concrete, since the concrete cannot be too cold or too hot. When the temperature decreases, the hydration reaction slows down. Hence, if the concrete temperature increases the reaction accelerates, creating an exothermic reaction -which itself produces more and more heat- causing temperature differentials within the concrete. This temperature gradient can lead to cracking. Moreover, during the initial phase of the life of the concrete, it is essential to avoid cracking caused by the rapid drying due to increased temperature and the on-going hydration reaction. The fourth section summarizes the results, the discussed issues, the advantage and disadvantages of the proposed SHM method.

2. Proposed SHM System

A. Description of Proposed SHM System, Components and Principles

In this paper, a SHM system with special characteristics was designed as shown in Fig. 1, then it is implemented and verified for a hypothetical bridge.

As shown in Fig. 1, parameters of temperature and humidity are monitored at two points of the bridge deck. This data will be used for data mining process and the prediction of critical values for the following days, and a warning system based on fuzzy inference techniques will assess the status of mentioned points and will announce timely pre-emptive alerts to the maintenance team of the bridges.

B. Description of Hardware and Wireless Sensor Nodes

Inexpensive and analog sensors of LM35 and HIH4000 were used for sensing and measuring the temperature and humidity respectively for the design of nodes of wireless sensor network. To assess the reliability and accuracy of the system, wired SHT11 sensor was used to obtain the temperature and humidity data in the desired nodes. This sensor calculates the humidity and temperature with high precision in digital form and does not need signal conditioning. A USB DAQ Digital Sensor was used to obtain its information. Sensor nodes were designed with Protel (Altium Designer) software. This design includes a board for LM35 and HIH4000 analog sensors and a separate board for SHT11 sensor which is considered as a reference for measurement. They are shown in Fig. 2 and Fig. 3, respectively.

In the next step, “ProBee-ZE10 ZigBee” module was used as wireless module. The default development board of this device was used to ensure easier application and also for easier installation of the connectors. Three selected wireless modules were configured. Two modules were defined as “end device” and one module was defined as “coordinator”. Configuration of end devices and coordinator was performed through USB terminal of a laptop and by the use of “Hyper Terminal” software.

3. Experimental Work

A. Negative Temperature Thermistor–Temperature Sensor

The first set of tests consisted of measuring the temperature with a NTC temperature sensor inside a concrete cube (common strength class C25/30, 10 cm length size). The acquisition system consists of a Sensor Board and an IRIS mote, facilitating the creation of an IEEE 802.15.4 network whose primary function is to remotely collect the data from the NTC sensor inside the concrete cube.

Fig. 2. layout of PCB of board and bias circuits of temperature and humidity sensors.

Fig. 3. The final layout of sensor nodes (Analog nodes including LM35 and HIH4000 sensors on the right side, and SHT node on the left side).
B. SHT15 Humidity and Temperature Sensor

In the second set of tests, the SHT15 digital sensor was used, facilitating to measure both temperature and humidity with high accuracy in a single chip sensor. Fig. 4 presents a schematic representation of process to measure the temperature and humidity within the concrete cube.

The conversion from the raw value returned by the SHT15 sensor, \( R_{\text{val}} \), to the temperature and humidity values was performed by using the following equations:

Temperature \[ \text{[} ^\circ \text{C} \text{]} \] = 0.01 \( R_{\text{val}} \) – 40 \hspace{1cm} (1)

Humidity \[ \text{[} \% \text{RH} \text{]} \] = –0.4 + 0.0405 \( R_{\text{val}} \) – 0.0000028 \( (R_{\text{val}})^2 \) \hspace{1cm} (2)

Before inserting the sensor inside the concrete block, the sensors were placed inside a small size cube (4 cm side length) made of cement mortar for its protection.

C. SHT21S Humidity and Temperature Sensor

1) Standalone Version

Besides the SHT15 sensor, we tested the new Sensirion SHT21S (humidity/temperature) sensor. Before testing this sensor a cement mortar shell has been used for its protection. This sensor is an updated version of the previous one but with a smaller package. To test this sensor, an acquisition system was designed to facilitate the acquisition of the analogue signal while converting it for its digital representation. As previously mentioned we intend to measure both temperature and humidity inside the concrete block, from the early ages, during setting and hardening period. The temperature and humidity values are obtained by using Eqs. (3) and (4) respectively:

Temperature \[ \text{[} ^\circ \text{C} \text{]} \] = –46.85 + 175.72 \times \frac{V_{\text{SO}}}{V_{\text{DD}}} \hspace{1cm} (3)

Humidity \[ \text{[} \% \text{RH} \text{]} \] = –6 + 125 \times \frac{V_{\text{SO}}}{V_{\text{DD}}} \hspace{1cm} (4)

where \( V_{\text{DD}} \) is the supply voltage at which the SHT21S sensor works, as presented in the datasheet of the sensor in the interface specifications. In this case, \( V_{\text{DD}} = 3 \text{ V} \).

Besides, since the SHT21S output is a Sigma Delta Modulated (SDM) signal, normally this signal is converted to an analogue voltage signal by the means of a low-pass filter. The output of low pass filter provides a voltage value \( (V_{\text{SO}}) \) which is a portion of \( V_{\text{DD}} \), depending on the measured humidity or temperature. The developed acquisition system (for the standalone SHT21S) incorporates a micro SD module, responsible for storing the values acquired from the SHT21S sensor, as shown in Fig. 5.

The “MSP430F449-STK2” module was used to convert the signal output from the RC-filter to the digital format. The algorithm running inside the microcontroller performs five readings (with a 100 ms interval between two consecutive readings for the temperature), storing the fifth reading in a buffer. Then, it switches to the humidity sensor, performing another five readings and conversions with the same duration between consecutive readings, storing the fifth reading in another buffer. Finally, after that it sends the commands to store the temperature and humidity values in separated text files, into the micro SD card.

2) Wireless Version

The SHT21S wireless prototype aims at creating a Building Wireless Sensor Network (BWSN) capable of measuring temperature and humidity inside a concrete structure. It has two Integrated Circuits (ICs) interfaces via Serial Port Interface (SPI), and an antenna allowing for connectivity with no additional hardware components. Besides, it provides real-time data information and remote interaction with multiple devices (e.g., laptop, PDA, cell phone with ZigBee capabilities). The “MSP430F2274” ultra-low-power microcontroller controls the “CC2500” radio transceiver-that operate at the 2.4 GHz band-and establishes a basic wireless network with minimal power requirements, enabling to extend the system lifetime. Fig. 6 presents the acquisition system used to read the signal from the SHT21S sensor. The computed temperature and humidity values are sent wirelessly to the access point. The end device reports periodically values each minute to the AP. The user depending on the application scenario can change this reporting periodicity value.

---

Fig. 4. Schematic representation of temperature and humidity sensor inside a concrete cube.

Fig. 5. SHT21S acquisition system for the standalone version.
D. Joint Verification of Shielded SHT15 and SHT21S Sensors

The main purpose of shielding the SHT15 and SHT21S sensors is to protect the sensor from the concrete high relatively humidity alkaline environment that could affect the sensor inside the concrete. Besides, the unique capacitive sensor element used to measure humidity as well as the band-gap sensor utilized to measure the temperature do not resist to the high relative humidity alkaline environment present in cement. To overcome this limitation, in the second series of tests we have decided to use a filter cap allowing for protecting the SHT15 and SHT21S humidity and temperature sensors against dust, water immersion, condensation, as well as contamination by particles. The cavity inside the filter cap is made such that the volume between the membrane and the sensor is kept minimal, which reduces the impact on the response time for the humidity measurements. Mounting schematics of the filter cap protection for the SHT15 and SHT21S sensors is shown in Fig. 7.

4. Results and Discussion

A. Evaluation of Real-time Monitoring of Temperature and Humidity

SHM related parameters can be assessed based on the humidity and temperature values of the proposed monitoring system. As previously mentioned in the introduction section, temperature and humidity can cause damage to the bridge structure including cracking caused by temperature gradient- which itself is caused by different degrees of sunlight on different parts of the bridge-, corrosion caused by humidity and climatic factors- e.g. corrosive sea salt- and those corrosions and damages that have different origin, but in which humidity and temperature act as an accelerating factor. Assessment and comparison of temperature and humidity values with critical threshold values at different points provides the possibility of detecting present structural issues or those that are going to happen.

To assess the accuracy and reliability of the proposed system, it was deployed for 3361 minutes-approximately two and a half days- to store temperature and humidity data then results were compared with the results of SHT11 sensor which its digital temperature and humidity data was collected through a data acquisition card via wired connection. The results can be stored the databases as shown in Table 1 and Table 2. They can also be exported to the other software -like MATLAB- for further processing. Fig. 8 shows the temperature data that were stored by wireless monitoring, and Fig. 9 shows the temperature data that were stored by wired SHT11 sensor in the same period. A similar process was also used to store humidity data. Humidity data logged by wireless sensor network monitoring system and HIH400 analog sensor is shown in Fig. 10 and humidity data logged by wired SHT11 digital sensor is shown in Fig. 11. Relative error of proposed SHM system in the calculation of temperature and humidity is shown in Fig. 12.
Table 1. An example of temperature values logged in the Excel files (TEMP1, TEMP2).

<table>
<thead>
<tr>
<th>DATE</th>
<th>TIME</th>
<th>TEMPERATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>04/14/2015</td>
<td>5:18 pm</td>
<td>29.7°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:19 pm</td>
<td>27.7°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:20 pm</td>
<td>27.5°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:21 pm</td>
<td>27.5°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:22 pm</td>
<td>27.5°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:23 pm</td>
<td>27.5°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:24 pm</td>
<td>27.5°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:25 pm</td>
<td>27.5°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:26 pm</td>
<td>27.5°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:27 pm</td>
<td>27.5°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:28 pm</td>
<td>27.4°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:29 pm</td>
<td>27.5°C</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:30 pm</td>
<td>28.0°C</td>
</tr>
</tbody>
</table>

Table 2. An example of humidity values logged in the Excel files (HUM1, HUM2).

<table>
<thead>
<tr>
<th>DATE</th>
<th>TIME</th>
<th>HUMIDITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>04/14/2015</td>
<td>5:18 pm</td>
<td>35.0%</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:19 pm</td>
<td>36.0%</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:20 pm</td>
<td>36.3%</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:21 pm</td>
<td>36.3%</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:22 pm</td>
<td>36.3%</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:23 pm</td>
<td>36.3%</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:24 pm</td>
<td>36.3%</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:25 pm</td>
<td>36.3%</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:26 pm</td>
<td>36.3%</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:27 pm</td>
<td>36.3%</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:28 pm</td>
<td>35.0%</td>
</tr>
<tr>
<td>04/14/2015</td>
<td>5:29 pm</td>
<td>36.5%</td>
</tr>
</tbody>
</table>

\[
\sigma(y) = \beta E_0 e^{-y} - \nu y
\]

\[
\varepsilon_T = \varepsilon_y - \varepsilon_T
\]
In the above equations, $\beta$ is a dimensionless function of $X$ or position along the bridge model. Values of $\beta$ are in the range between zero- which indicates to the full damage and total loss of EL- and one –which indicates to the status without any damage and one hundred percent intact-. When the value of $\beta$ function is known, thermal response of a damaged bridge can be estimated and calculated by Eq. (5) and Eq. (6) as shown in Fig. 13. More comprehensive information regarding the operations and calculations related to the use of the thermal response of the structure for its SHM is available in [28]-[30].

B. Negative Temperature Thermistor – Temperature Sensor

The first experimental approach for reading temperature inside a concrete cube involves the use of a negative temperature coefficient thermistor and an IRIS mote. This setup foresees an automatic wireless monitoring system. The temperatures inside the concrete cube and environment have been compared. There is a difference of 5º C between the actual and measured temperatures. This is due to some failures during the calibration of the sensor, resulting in inaccurate values. Based on this fact, we can conclude that using a NTC sensor with an “unknown” behavior is not the most adequate approach to the problem. Besides, this kind of sensor is not able to simultaneously measure temperature and humidity inside the concrete structures.

C. SHT15 Humidity and Temperature Sensor

The second set of tests considers the use of the SHT15 sensor, allowing for measuring both the humidity and temperature. Two solutions were tested, one with the PIC18F4680 microcontroller and another one using the Arduino platform. Before using the SHT15 sensor in a real scenario, some tests have been performed to verify the accuracy of the temperature and humidity readings. To measure humidity we place the SHT15 sensor inside a small mortar cube for sensor protection. When the cubes were placed in a tray (with 2–3 mm water level), we observed the rise of water inside the cube by capillary. After around one minute, the humidity reaches a value of 98% RH. The objective of this test was to verify the sensor integrity, as well as the porosity effect of its mortar shell. The results obtained from both PIC18F4680 and Arduino platform were identical. The tests were carried out during several hours, to observe if any variation of humidity and temperature could be detected.

In another experiment a SHT15 sensor with a mortar shell was fully immersed in water. One observes that the temperature was decreasing while the humidity was increasing. After 20 min of accurate measurements, we have decided to prolong the test during one week. However, after one day, the SHT15 temperature sensor went off. Then, after 4 days the same happened to the humidity sensor. It is believed that the primary reason for this occurrence is that some chemical reactions inside the mortar shell have affected the capacitance of the sensor. Sensor components might not resist to alkaline ions present in cement, namely calcium hydroxide, which can be released in water from its mortar shell during immersion. To solve this problem, instead of making a cement-based mortar shell it may be preferable to shield the sensor using other material, textile or polymer based.

D. SHT21S Humidity and Temperature Sensor

1) Standalone Version

The SHT21S sensor protected by a mortar cube was placed inside a concrete cube during testing. The values measured by the sensor were recorded into the micro SD card. The measurements were performed in outdoor environmental conditions during summer. During the first 12 h there has been a constant and progressive variation; while between the 12th and 16th hours a decrease in the temperature and humidity values has been observed. After 16 h, the sensor stopped reading the temperature values. Only the humidity values have been measured beyond this time instant. As occurred with SHT15, the SHT21S sensor components have not resisted long time inside the concrete alkaline environment. To overcome this limitation, shielding of sensor is also advised in this case, e.g., with textile, polymer Poly-Butylene-Terephthalate or even metal shielding.

2) Wireless Version

In scenarios of remote monitoring, there is a need of extracting and recording the data gathered by the sensor nodes. To avoid the need of regularly visiting, remote access to the collected data is essential. Moreover, solutions involving WSNs have a tremendous potential in real time structural health monitoring, since they potentially reduce costs.

The SHT21S wireless version allows for collecting the information from any given structure via the SHT21S ez430-RF2500 C++ software program which is responsible for the acquisition of the values from the SHT21S sensor. To analyze the acquired values we can export the data to a MATLAB-file. The results obtained for temperature and humidity are quite accurate. Therefore, the use of a porous cement mortar as protective shell does not affect the sensor readings. This method of protection of the sensor is similar to those developed by [27] which is recently published, although unknown to authors during the experimental phase. However, the presented solution does not consider an encapsulation box for the electronic acquisition system components -since it is outside the “brick”-. This way we are able to obtain more accurately values for the temperature and humidity since the sensor is placed as close to the environment as possible. By using an encapsulation box the detected temperature and humidity may not be the actual structure temperature and humidity, as stated in [28]. Besides, in the work developed by [27] the Radio Frequency Integrated Circuit (RFIC) transmitter is inside the brick, being the maximum effective reception range bellow 20 m.

The research also considers a package to protect the sensor from the aggressive environment. Preliminary results show that the transmission distance is strongly affected by the steel backed formwork, showing that the maximum distance achieved without the formwork is 7.5 m. In our case, by considering the open field scenario -since the acquisition systems is outside the “brick”- the ez430-RF2500 can achieve a minimum effective reception range
of 35 m. In this experiment, the SHT21S sensor temperature readings have been successfully performed during the first 16 h, while the humidity values were successfully obtained during the first 21 h. After this period, the sensor went off which is possibly caused by the alkaline concrete environment that stopped the sensor operation.

E. Joint Verification of Shielded SHT15 and SHT21S Sensors

In this set of experiments, the SHT15 and SHT21S sensors, which are previously shielded, were inserted into two small mortar cubes -4 cm of side length- before being inserted into the concrete block. To test the accuracy of the measurements, the mortar cubes were first placed in a tray with water. After some time, we observed the rise of water inside the cube by capillarity. Then, the cubes were removed for drying. As is shown in Fig. 14, the cube containing the SHT15 sensor initiated the drying process after 10 h, while the cube containing the SHT21S sensor started the drying process after 60 h. After 97 h, we have repeated the test of placing the cubes in a tray with water, in order to observe the increase of the humidity values.

The standard deviation between the humidity values, which are measured by the SHT15 sensor and the SHT21 sensor, is explained by the fact that the small cubes are not exactly the same, so some variations in terms of humidity may exist during the drying process. Also, if the SHT21S is exposed to conditions outside the normal operation range (humidity > 80%), an offset could exist. Therefore, in high relative humidity environment it is advised to use the SHT15 sensor. The measured temperature is similar to the ambient temperature. As can be seen in circle 1 in Fig. 14 in the first 20 hours of testing, both SHT15 sensor and SHT21s sensor measure the temperature carefully with appropriate adaptation in response to the changes in ambient temperature, but in almost 22th hours of the start of the experiment, both of them sense a sharp drop in temperature. With increase of ambient temperature, they measure high temperatures in steady state which represents the appropriate sensitivity in the application of structural health monitoring of concrete structures. On the other hand, this represents the stability and performance of both sensors within the first 25 hours of test.

Regarding to the circle 2 in Fig. 14, it should be considered at the same constant ambient temperature, both sensors reported about 0.5 °C temperature swings that the interpretation of which can be differences in the type of cement cube which is used to protect them. The cubes are not exactly in the same structure and geometry so some differences and swings occur in the above mentioned range in respect to the reference temperature. Regarding to the circles 3 and 4 in Fig. 14, there are similar explanation about measuring of the temperature which have been described before. They implicitly refer to the failure of protecting shielded system using concrete cubes which caused some inertia and errors in temperature measurements by both of above mentioned sensors.

Regarding to the circle 5 in Fig. 14, after starting the drying process, the reference sensor is broken but SHT15 and SHT21s sensors continue to operate. This indicates the following two points:

- Capability of both sensors to measure temperature within or adjacent concrete.
- High performance of the proposed protection system including concrete cubes.

Regarding to the rectangle 6 in Fig. 14, as previously mentioned, it should be pointed out about the lack of proper functioning of the SHT21s humidity sensor in the first 60th hours of the test which is in the saturated conditions. This part of the test is reached us to the following conclusions:

- Non-performance of the protection system, including concrete cubes for protection of humidity sensor SHT21s.
- The usage of the humidity sensor SHT15 in SHM applications at environments with high moisture (more than 90%), is preferable.

In the final step of this study, the SHT15 and SHT21S sensors, which were previously shielded and protected by a mortar cube, were placed inside a concrete cube during the test. The values measured by the sensors were recorded into the micro SD card. The data collected from the sensors is shown in Fig. 15. The tests

![Fig. 14. Results for the humidity and temperature obtained using the SHT15 and SHT21S sensors during 5 days.](image-url)
were performed during 6 days. During the first 12 h, there is an increase of the humidity for the cube containing the SHT15 sensor, while the humidity inside the cube containing the SHT21S sensor achieves the maximum value (i.e., 100%). During the curing process, the temperature inside the cubes was about 37 °C after the first 11 h. Therefore, we may conclude that by using a filter cap as shown in Fig. 7, we protect the humidity and temperature sensors and a long-term solution for SHM is obtained. This is one of the major goals in this study. Finally, it was verified that both sensors were performing measurements inside the concrete after 2 months of experiments.

In fact Fig. 15 is laboratory-controlled version of the early stages of the formation of concrete and hydration process. Circle 1 in Fig. 15 represents the initial phase of the formation of concrete where the temperature has gone up- which is induced by rapid hydration rate- while high level of humidity still remains. Under these conditions, both temperature and humidity sensors SHT15 and SHT21S have good performances. After the initial phase of hydration, by reducing the temperature and humidity during a process which is known as “drying”, SHT21S remains in the saturated condition but SHT15 moisture sensors continues to measure the humidity well. This demonstrates the capabilities of SHT15 in the real-time & continuous measurements of humidity in the vicinity of severe alkaline environment of concrete. On the other hand, by comparison of the SHT15 results with reference sensor results, we found that organic material protection system does not affect the accuracy and sensitivity of the SHT15. It helps SHT15 for withstanding and measuring of the temperature and humidity with high acceptable accuracy which is shown in the area inside the circle 3 in Fig. 15.

Throughout the mentioned period, both sensors measure the temperature with reasonable accuracy. This is the evidence of this fact that the shielding protection system has not any bad effects on the accuracy of the temperature measurements. It increases the lifespan of temperature sensors via shielding organic materials.

In the Circle 3 in Fig. 15, it can be seen some differences in the results of temperature sensors SHT15 and SHT21S. The reason of those differences, as we have already mentioned, is that the two sensors did not exposure on equal conditions within the desiccator. Another reason is the dissimilarity in the types of the shielding which are not 100% equal and have slightly different properties.

Regarding to the rectangle 4 in Fig. 15, during all of this period, humidity sensor SHT21S is in saturation mode and does not show an appropriate performance in measuring of the humidity in the presence of an alkaline environment in the concrete. On the other hand, this behavior reflects the inefficiency of the proposed shielding protection system in the protection of the humidity sensor SHT21S.

As a summary of Fig. 15, which is in fact a controlled-laboratory conditions to investigate the processes of hydration and the initial formation of the concrete, it should be noted that the system including a sensor SHT15 and proposed shielding protection system (as stated before in section 3.4), has an acceptable functionality which is required for being used in SHM of the concrete structures.

5. Conclusion

In this study, a system with multiple functions for monitoring the structural health of a medium sized assumed bridge with the use of wireless sensor networks was designed, implemented, and simulated. The proposed SHM system monitors the parameters of temperature and humidity in two points of the bridge deck. A wireless sensor network has been created based on the IEEE 802.15.4, allowing for the creation of a continuous monitoring system capable of sending data wirelessly. In this network, the nodes used two types of sensors, the SHT15 and SHT21S ones, to read both humidity and temperature, in real-time and continuous monitoring basis. The obtained results show two types of sensors and the measurement procedure have highly potential for inexpensive concrete structure monitoring. However, during the first set of experiments the SHT15 temperature
sensor stopped working after one day. Besides, after four days, the same happened to the humidity sensor. The temperature readings from the SHT21S sensor have been successfully performed during the first 16 h of the experiment, while the humidity values were successfully obtained for the first 21 h. After this period the sensor went off. The initial sets of results were very promising, although SHT15 and SHT21S sensors went off after some time inside concrete. This is explained by the fact that the components of the sensors do not resist to concrete high relative humidity alkaline environment.

The experiments carried on also shown that a porous cement mortar could be used as shell to protect sensor wire connections. High porosity of this mortar shell easily allows moisture and temperature measures of involving concrete but this solution does not protect sensors of the alkaline environment. Hence, it is advised to shield the sensors before using them within the structure, especially during the construction phase. It is also advised to use sensors that can resist in concrete alkaline environment.

In the end, given the accuracy and reliability of assessments and analysis results and the much lower costs of this system in terms of initial equipment and maintenance (due to the simplicity of its structure) proposed SHM system can be used for the long-term and real-time monitoring of medium-sized to large-sized reinforced concrete bridges.

References


Saman Shoorabi Sani received the B.S. and M.Sc. degrees in electrical engineering from Hakim Sabzevari University in 2009 and 2015, respectively. His current research includes analog/RF integrated circuit design and wireless sensing.

Majid Baghaei-Nejad received the B.S. and the M.Sc. degrees in electrical engineering from Ferdowsi University of Mashhad and Tarbiat Modares University, Iran, in 1996 and 2000, respectively, and his Ph.D. degree in electronic and computer systems from Royal Institute of Technology (KTH), Stockholm, Sweden in 2008. Since 2000 he has been with electronics department at Hakim Sabzevari University, Iran, where currently has an assistance professor position. Also, he has been there the Dean of the Faculty of Electrical and Computer Engineering since 2014. His current research includes low power analog/RF integrated circuit design, ultra wideband communication, RFID systems and wireless sensing.

Mona Kalate Arabi received the B.S. and the M.Sc. degrees in electrical engineering from Islamic Azad University of Sabzevar and Islamic Azad University of Neyshabur, Science & Research Branch in 2009 and 2015, respectively. Her current research includes analog integrated circuit design and remote sensing.
A Novel Approach to Speaker Weight Estimation Using a Fusion of the i-vector and NFA Frameworks

Amir Hossein Poorjam, Mohamad Hasan Bahari, and Hugo Van hamme

Abstract. This paper proposes a novel approach for automatic speaker weight estimation from spontaneous telephone speech signals. In this method, each utterance is modeled using the i-vector framework which is based on the factor analysis on Gaussian Mixture Model (GMM) mean super vectors, and the Non-negative Factor Analysis (NFA) framework which is based on a constrained factor analysis on GMM weight super vectors. Then, the available information in both Gaussian means and Gaussian weights is exploited through a feature-level fusion of the i-vectors and the NFA vectors. Finally, a least-squares support vector regression (LSSVR) is employed to estimate the weight of speakers from the given utterances.

The proposed approach is evaluated on spontaneous telephone speech signals of National Institute of Standards and Technology (NIST) 2008 and 2010 Speaker Recognition Evaluation (SRE) corpora. To investigate the effectiveness of the proposed approach, this method is compared to the i-vector-based speaker weight estimation and an alternative fusion scheme, namely the score-level fusion. Experimental results over 2339 utterances show that the correlation coefficients between the actual and the estimated weights of female and male speakers are 0.49 and 0.56, respectively, which indicate the effectiveness of the proposed method in speaker weight estimation.

Keywords: i-vector, non-negative factor analysis, least-squares support vector regression, speaker weight estimation.

1. Introduction

The voice of a speaker conveys information about speaker’s traits and states such as age, gender, body size (weight/height) and emotional state. Estimation of speaker’s weight (which is considered as a long term trait of a speaker and an important parameter in various applications) is an interesting and challenging task in forensic, medical and commercial applications. In forensic scenarios, body size estimation of suspects from their voices can direct investigations to find cues in judicial cases. In service customization, automatic weight estimation may help users to receive services proportional to their physical conditions.

The relation between the size of various components of the sound production system (such as vocal folds and vocal tract) and the body size of a speaker has motivated researchers in the field of speaker recognition to look for features of an acoustic signal that provide cues to the body size of speakers. For instance, authors in [1] found a relationship between formants and the length of the vocal tract, based on the source-filter theory. Thus, since the vocal tract is a part of speaker’s body, this feature can be used to estimate the weight of a speaker [2].

However, speaker weight estimation from the voice patterns is challenging. For instance, mean fundamental frequency (f0) of voice is reported as a feature which has a (negative) correlation with body size. That is, females and children have higher f0, while in males (who are taller and heavier), this value is lower [3]. However, when the relation of the fundamental frequency (f0) and weight was investigated within male and female speakers, no correlation was found between f0 and the weight of adult humans [4], [5].

The lowest fundamental frequency of voice (F0 min) is another feature which is determined by the mass and length of the vocal folds [3]. By investigating this feature, researchers have found no correlation between F0 min and weight in adult human speakers [4], [5].

Fitch has found formant dispersion (the averaged difference between adjacent pair of formant frequencies) a reliable feature which has a correlation with both vocal tract length and body size in macaques [6]. However, a weak relation between formant parameters and weight of human adults is reported in study conducted by Gonzalez [7]. This weak correlation may be due to the fact that the vocal folds in humans at puberty grow independent of the rest of the head and body. This issue is more evident in the males than the females [8], [9].

Gonzalez studied the correlation between formant frequencies and weight in human adults [7]. He calculated the formant parameters by means of a long-term average analysis of running speech signals uttered by 91 speakers. In this experiment, the Pearson correlation coefficients between formants and weights for male and female speakers were reported to be 0.33 and 0.34, respectively [7].

In research conducted by Van Dommelen and Moxness [10], the ability to judge the weight of speakers from their speech samples was investigated. They reported a significant correlation between estimated weight (judged by listeners) and actual weight of only male speakers. In addition, they performed a regression analysis involving several acoustic features such as f0, formant frequencies, energy below 1 kHz, and speech rate. The results showed...
that the speech rate was the only parameter which had a significant correlation with male speaker’s weight. They concluded that speech rate of male speakers is a reliable predictor for weight estimation.

Modeling speech utterances with Gaussian mixture model (GMM) mean super vectors is demonstrated to be an effective approach to speaker recognition [11]. However, GMM mean super vectors are high dimensional vectors, and obtaining a reliable model is difficult when limited data are available.

Recently, utterance modeling using the i-vector framework [12] has considerably increased the accuracy of the classification and regression problems in the field of speaker characterization [13]-[15]. The i-vector, which is based on the factor analysis on GMM mean super vectors, represents an utterance in a compact and a low-dimensional feature vector. In addition, various studies show that although GMM weights convey less information than GMM means, they provide complementary information to GMM means [16]-[18]. A Non-negative Factor Analysis (NFA) framework [16], which is based on a constrained factor analysis for GMM weights, has been recently introduced and yields a new low-dimensional utterance representation. In [19] we successfully applied a score-level fusion of the i-vector and the NFA frameworks to simultaneously estimate various characteristics of speakers from speech signals. We showed that utilizing information in both GMM means and GMM weights through a score-level fusion of the i-vector and the NFA frameworks is an effective approach to improve the estimation accuracy. However, the fusion at score level requires a large development data set to train the fusion model, which results in decreasing the number of available training data.

In this study, a new speech-based method for automatic weight estimation is proposed in which instead of using raw acoustic features, each utterance is modeled using a fusion of the i-vector and the NFA frameworks at feature level. In this new utterance modeling approach, in addition to exploiting the available information in GMM means and GMM weights, the need for assigning a considerable amount of training data for development set is eliminated and speaker weight estimation is performed in one learning phase. To perform function approximation, a least-squares support vector regression (LSSVR) is utilized in this paper.

For the comparison purpose, the proposed method is compared to the i-vector-based speaker weight estimation and speaker weight estimation using an alternative fusion scheme, namely the score-level fusion. The proposed approach is evaluated on spontaneous telephone speech signals of the NIST 2008 and 2010 SRE corpora. Experimental results confirm the effectiveness of the proposed approach in automatic speaker weight estimation.

The rest of the paper is organized as follows. In Section 2 the problem of automatic weight estimation is formulated and different baseline systems for speaker weight estimation are described. The proposed approach is elaborated in Section 3. Section 4 explains the experimental setup. The evaluation results are presented and discussed in Section 5. The paper ends with conclusions in Section 6.

2. Automatic Weight Estimation from Speech Signals

In this section, the problem of automatic weight estimation is formulated and different baseline approaches are described.

A. Problem Formulation

In the speaker weight estimation problem, we are given a set of training data \( D = \{O_i, y_i\}_{i=1}^{N} \), where \( O_i \) denotes the \( i \)-th utterance and \( y_i \in \mathbb{R} \) denotes the corresponding weight.

The goal is to approximate a function \( g \), such that for an utterance of an unseen speaker, \( O_{un} \), the estimated weight, \( \hat{y} = g(O_{un}) \), approximates the actual weight as good as possible.

B. Baseline Approaches

For the comparison purpose, the proposed method is compared to three baseline approaches, namely the basic estimator, the i-vector-based speaker weight estimation [17] and speaker weight estimation using score-level fusion of the i-vector and the NFA frameworks [19].

1) Basic Estimation System:

The output of a basic estimator is the average weight of speakers in training data set. The basic estimation system provides us a chance level accuracy.

2) The i-vector-based System:

In the i-vector-based weight estimation system, each utterance is mapped onto a 400 dimensional vector using the i-vector framework. Then, the extracted i-vectors along with their corresponding weight labels are used to train estimator. This method is considered as the baseline system in this paper.

3) The Score-level Fusion System:

The fusion of the i-vector and the NFA frameworks is an effective approach to exploit the available information in both GMM means and GMM weights and consequently to enhance the estimation accuracy. In this method, each utterance is converted to an i-vector and an NFA vector. Then, the obtained vectors of the train set are employed to train the i-vector-based and the NFA-based models. In the next step, the i-vectors and the NFA vectors of development set are applied to the trained models. The outputs are then concatenated to form 2-dimensional vectors and along with the corresponding weight labels are used to train the fusion model. This system, which is labeled as the score-level fusion system in this paper, is presented to investigate the effect of different fusion schemes in speaker weight estimation.

3. System Description

A. Utterance Modeling

By fitting a GMM to acoustic features extracted from each speech signal, a variable-duration speech signal is converted into a fixed-dimensional vector which is suitable for regression algorithms. The parameters of the obtained GMM characterize the corresponding utterance. Due to
limited data, we are not able to accurately fit a separate
GMM for a short utterance, especially in the case of GMMs
with a high number of Gaussian components. Thus, for
adapting a universal background model (UBM) to
characteristics of utterances in training and testing
databases, parametric utterance adaptation techniques are
applied. In this paper, the i-vector and the NFA frameworks
are applied to adapt UBM means and weights, respectively.

1) Universal Background Model and Adaptation:
Consider a UBM with the following likelihood function
of data \( O = \{ o_1, \ldots, o_T \} \).

\[
p(o_t | \gamma) = \sum_{c=1}^{C} \pi_c p(o_t | \mu_c, \Sigma_c)
\]
\[
\gamma = \{ \pi_c, \mu_c, \Sigma_c \}, c = 1, \ldots, C
\]
(1)

where \( o_t \) is the acoustic vector at time \( t \), \( \pi_c \) is the mixture
weight for the \( c^{th} \) mixture component, \( p(o_t | \mu_c, \Sigma_c) \) is a
Gaussian probability density function with mean \( \mu_c \) and
covariance matrix \( \Sigma_c \), and \( C \) is the total number of Gaussian
components in the mixture. The parameters of the UBM –
\( \gamma \) – are estimated on a large amount of training data.

2) The i-vector Framework:
One effective method for speaker weight estimation involves adapting UBM means to the speech characteristics
of the utterance. Then, the adapted GMM means are
extracted and concatenated to form Gaussian mean super
vectors. However, since Gaussian components of the UBM
model are adapted independent of each other, some
components are not updated in the case of limited training
samples [20]. This problem can be alleviated by linking the
Gaussian components together using the Joint Factor
Analysis (JFA) framework [21].

In the JFA framework, each utterance is represented by a
super vector \( \mathbf{M} \) which is a speaker- and channel-dependent
vector of dimension \( C.F \), where \( C \) is the total number
of the mixture components in a feature space of dimension \( F \).
In the JFA framework, it is assumed that \( \mathbf{M} \) can be
decomposed into two super vectors:

\[
\mathbf{M} = \mathbf{s} + \mathbf{c}
\]
(2)

where \( \mathbf{s} = \mathbf{u} + \mathbf{Vq} + \mathbf{Dr} \) is a speaker-dependent super vector
and \( \mathbf{c} = \mathbf{Up} \) is a channel-dependent super vector. \( \mathbf{s} \) and \( \mathbf{c} \) are independent and possess normal distributions.
\( \mathbf{u} \) is the speaker- and channel-independent super vector, \( \mathbf{V} \) defines a
lower dimensional speaker subspace, \( \mathbf{U} \) is a lower
dimensional channel subspace, and \( \mathbf{D} \) defines a speaker
subspace. \( \mathbf{q} \) and \( \mathbf{r} \) are factors in speaker subspace, and \( \mathbf{p} \) is a
channel-dependent factor in channel subspace. The vectors
\( \mathbf{p}, \mathbf{q} \) and \( \mathbf{r} \) are random variables with standard normal
distributions \( \mathcal{N}(0, I) \) which are jointly estimated.

In the JFA framework, the channel factor contains some
information about speakers, which can be utilized in
speaker identification. This fact resulted in proposing a new
utterance modeling approach, referred to as the i-vector
framework or the total variability modeling [22]. This
method comprises both speaker variability and channel
variability. Channel compensation procedures such as
within-class covariance normalization (WCCN) can be
further applied to compensate the residual channel effects
in the speaker factor space [23].

The i-vector framework assumes that each utterance
possesses a speaker- and channel-dependent GMM super
vector which it’s mean, \( \mathbf{M} \), can be decomposed as

\[
\mathbf{M} = \mathbf{u} + \mathbf{Tv}
\]
(3)

where \( \mathbf{u} \) is the mean super vector of the UBM, and \( \mathbf{T} \) spans
a low-dimensional subspace (400 dimensions in this work).
In the i-vector framework, \( \mathbf{T} \) and \( \mathbf{v} \) are estimated using the
Expectation-Maximization (EM) algorithm. In the E-step, \( \mathbf{T} \)
is supposed to be known, and \( \mathbf{v} \) is updated. In the M-step, \( \mathbf{v} \)
is assumed to be known, and \( \mathbf{T} \) is updated. The subspace vector \( \mathbf{v} \) is treated as a hidden variable with the standard
normal prior and the i-vector is its maximum-a-posteriori
(MAP) point estimate which is calculated by maximization
of the following auxiliary function over \( \mathbf{v} \).

\[
\Psi(\gamma, \mathbf{v}) = \sum_{c=1}^{C} \sum_{t=1}^{T} \theta_{c,t} \log \pi_c p(o_t | [\mu_c + T_v], \Sigma_c) \mathcal{N}(\mathbf{v})
\]
(4)

where \( \mathcal{N}(\mathbf{v}) \) denotes the standard normal distribution of \( \mathbf{v} \),
\( \mathbf{T} \) are the rows of the subspace matrix \( \mathbf{T} \), which correspond
to the \( c^{th} \) Gaussian mean, and \( \theta_{c,t} \) is the occupation count for
the \( c^{th} \) mixture component and \( t^{th} \) frame. The occupation
count is calculated as follows:

\[
\theta_{c,t} = \frac{\pi_c p(o_t | \mu_c, \Sigma_c)}{\sum_{c=1}^{C} \pi_c p(o_t | \mu_c, \Sigma_c)}
\]
(5)

In the E-step, the posterior distribution of \( \mathbf{v} \) is Gaussian
with the following mean \( \mathbf{v}_\mu \) and covariance matrices \( \mathbf{v}_\sigma \)
[24]:

\[
\mathbf{v}_\mu = \mathbf{v}_\mu + \mathbf{T} \Sigma^{-1} \mathbf{T}^\top \theta_{c,t} (o_t - \mathbf{m}_c)
\]
(6)

\[
\mathbf{v}_\sigma = \mathbf{I} + \sum_{c=1}^{C} \theta_{c,t} \Sigma^{-1} \mathbf{T} \Sigma^{-1} \mathbf{T}^\top \theta_{c,t}
\]
(7)

where \( \mathbf{I} \) denotes an identity matrix of appropriate size, \( \mathbf{m}_c \),
and \( \Sigma_c \) are adapted mean and covariance of the \( c^{th} \)
Gaussian, which are updated during each EM iteration
starting from UBM parameters, and \( \mathbf{T} \) represents the
transpose operator.

In the M-step, the subspace matrix \( \mathbf{T} \) is estimated via
maximization of the following auxiliary function over \( \mathbf{T} \):

\[
\Psi(\gamma, \mathbf{T}) = \sum_{c=1}^{C} \sum_{t=1}^{T} \theta_{c,t} \log \pi_c p(o_t | [\mu_c + T_v], \Sigma_c)
\]
(8)

An efficient procedure for training \( \mathbf{T} \) and for MAP
adaptation of the i-vectors can be found in [24].
In the total variability modeling approach, the i-vector is the low-dimensional representation of an audio recording that can be used for classification and estimation purposes.

3) The Non-negative Factor Analysis (NFA) Framework:

The NFA is a new framework for adaptation and decomposition of GMM weights based on a constrained factor analysis [16]. The basic assumption of this method is that for a given utterance, the adapted GMM weight super vector can be decomposed as follows:

\[ w = \pi + Lr, \]  

where \( \pi \) is the UBM weight super vector (2048 dimensional vector in this study), \( L \) is a matrix of dimension \( C \times \rho \) spanning a low-dimensional subspace (300 dimensions in this work), and \( r \) is a low-dimensional subspace vector obtained through a constrained maximum likelihood estimation criterion.

In this framework, the adapted weights are obtained by maximizing the following objective function over \( w_c \):

\[
\Psi(y,r) = \sum_{c=1}^{C} \sum_{i=1}^{n} \theta_{ci} \log w_e(o_i|\mu,\Sigma_e)
\]

Substituting \( w_e \) by \( (\pi + Lr) \) in the Eq. 10, and given an utterance \( O \), a maximum likelihood estimation of \( r \) is obtained by solving the following constrained optimization problem:

\[
\begin{aligned}
\max_r (\Psi(y,r)) &= \max_r (\tilde{\theta}(O)\log(\pi + Lr)) \\
\text{Subject to} \quad &I(\pi + Lr) = 1 \\
&\pi + Lr > 0
\end{aligned}
\]

where \( I \) is a row vector of dimension \( C \) with all elements equal to one, and \( \tilde{\theta}(O) = \sum \{ \theta_{c1}, \ldots, \theta_{cn} \} \).

In this framework, neither the subspace matrix \( L \) nor the subspace vector \( r \) is constrained to be non-negative. However, unlike the i-vector framework, the applied factor analysis for estimating the subspace matrix \( L \) and the subspace vector \( r \) is constrained such that the adapted GMM weights are non-negative and sum up to one. The procedure of calculating \( L \) and \( r \) involves a two-stage algorithm similar to EM and can be found in [16]. The subspace matrix \( L \) is estimated over a large training dataset. It is then used to extract a subspace vector \( r \) for each utterance in train and test datasets.

This new low-dimensional utterance representation approach was successfully applied to speaker characterization [17], [19] and language/dialect recognition [16] tasks.

4) Feature-Level Fusion of the i-vectors and NFA Vectors:

Previous studies show that although GMM weight super vectors contain less information than GMM means, they provide complementary information to GMM means [18]. Feature-level fusion and score-level fusion are considered as effective approaches to exploit available information in both GMM means and weights [18], [19]. Score-level fusion, in which the outputs of different estimators are fused, requires a development data set to train the fusion model, which results in decreasing the number of training data. However, fusion at feature level, in which various features are normalized and concatenated, eliminates the need for assigning a considerable amount of available training data for development set, and estimation can be performed in one learning phase.

In this paper, a feature-level fusion of the i-vectors and the NFA vectors is considered to improve the estimation accuracy. As illustrated in Fig. 1, the extracted i-vectors and the NFA vectors are length normalized by having mapped onto a low-dimensional space using linear discriminant analysis (LDA) [25]. Then, the obtained low-dimensional vectors are concatenated to form a longer vector.

B. Function Approximation Using LSSVR

Support vector regression (SVR) is a function approximation approach developed as a regression version of the widely known Support Vector Machines (SVM) classifier. Using nonlinear transformations, SVMs map the input data onto a higher dimensional space in which a linear solution can be calculated. They also keep a subset of the samples which are the most relevant data for the solution and discard the rest. This makes the solution as sparse as possible. While SVMs perform the classification task by determining the maximum margin separation hyper plane between classes, SVR carries out the regression task by finding the optimal regression hyper plane in which most of training samples lie within an \( \varepsilon \)-margin around this hyper plane [26].

In this study, we use the least squares version of support vector regression. While an SVR solves a quadratic programming with linear inequality constraints, which results in high algorithmic complexity and memory requirement, an LSSVR involves solving a set of linear equations by considering equality constraints instead of inequalities for classical SVR [26], which speeds up the calculations.

In a regression problem, we are given a training dataset \( D^r = \{(o_1,y_1),\ldots,(o_n,y_n)\} \), where \( o_a \) and \( y_a \) denote a vector of observed features of the \( n^{th} \) data item and its corresponding output, respectively. The goal is to determine a function \( h(o) \) such that the outputs are predicted accurately. In primal form of LSSVR, \( h(o) \) is considered as

\[
h(o) = \beta \phi(o) + c
\]

Fig. 1. Block diagram of the utterance modeling in feature-level fusion.
A least squares loss function is applied instead of Vapnik’s\(\varepsilon\)-insensitive loss function in LSSVR to simplify the formulations to minimize
\[
\frac{1}{2}||\theta||^2 + \frac{1}{2} \beta \sum_{n=1}^{N} e^2_n
\]
subject to
\[
y_n = \theta \phi(o_n) + c + e_n
\]
where \(\beta\) is an error cost factor and \(e_n \in \mathbb{R}\) are error variables.

This optimization problem can be solved more efficiently for high dimensional data by using the Lagrangian variables \(\nu\) and minimizing the following dual cost function [26].
\[
\Omega(\theta, c, e, \nu) = \frac{1}{2}||\theta||^2 + \frac{1}{2} \beta \sum_{n=1}^{N} e^2_n
- \sum_{n=1}^{N} \nu_n [\theta \phi(o_n) + c + e_n - y_n]
\]
This minimization problem can be solved directly by taking the partial derivative of \(\Omega\) with respect to \(\theta, c, e\) and \(\nu\) and setting the results to zero. This results in solving a linear system of equations. Inserting the obtained results to (12) leads to the regression function
\[
h(o) = \sum_{n=1}^{N} \nu_n \langle \phi(o_n), \phi(o) \rangle + c = \sum_{n=1}^{N} \nu_n K(o_n, o) + c
\]
where \(K(o_n, o)\) denotes the kernel function and \(c\) and \(\nu\) are the solution to optimization problem (15).

One drawback of the applied simplification in LSSVR formulation is the loss of sparseness. Therefore, all samples contribute to the model, and consequently, the model often becomes unnecessarily large.

C. Training and Testing

The block diagram of the proposed weight estimation approach is shown in Fig. 2. During the training phase, each utterance in the training data set is mapped to a high dimensional vector using the feature-level fusion utterance modeling described in Section 3.A.4. Then, the obtained vectors along with their corresponding weight labels are used to train an estimator for approximating function \(g\). During the testing phase, the same utterance modeling approach applied in training phase is used to extract a high dimensional vector from a test utterance. Then, the estimated weight is obtained using the trained estimator.

4. Experimental Setup

A. Database

The National Institute for Standard and Technology (NIST) has held annual or biannual speaker recognition evaluations (SRE) for the past two decades. With each SRE, a large corpus of telephone (and more recently microphone) conversations is released. Conversations typically last 5 minutes and originate from a large number of speakers for whom additional meta-data is recorded.

The NIST databases were chosen for this work due to the large number of speakers and because the total variability subspace requires a considerable amount of development data for training. The development data set used to train the total variability subspace and UBM includes over 30,000 speech recordings and was sourced from the NIST 2004-2006 SRE databases, LDC releases of Switchboard 2 phase 3 and Switchboard Cellular (parts 1 and 2).

For the purpose of automatic speaker weight estimation, telephone recordings from the common protocols of the recent NIST 2008 and 2010 SRE databases are pooled together to create a dataset of 8241 utterances uttered by 1333 speakers. Then, it is divided to two disjoint parts such that 80% and 20% of all speakers are used for training and testing sets, respectively. Thus, of all 8241 utterances, 5902 utterances are considered for training set and 2339 utterances are considered for testing set. Fig. 3 shows the weight histograms of training and testing datasets for male and female speakers.

B. Performance Metric

In order to evaluate the effectiveness of the proposed method, the mean-absolute-error (MAE) of the estimated weight and the Pearson correlation coefficient (CC) between the actual and estimated weights are used. MAE is defined as:
\[
\text{MAE} = \frac{1}{N} \sum_{n=1}^{N} |w_n - \hat{w}_n|
\]
where \(w_n\) and \(\hat{w}_n\) are the actual and estimated weights, respectively.

Fig. 3. The weight histograms of telephone speech utterances of training and testing datasets for male and female speakers.
weight, and speaker weight estimation systems, the basic estimator row of Table 1. Besides providing a reference level for estimation using a basic estimator are reported in the first row of Table 1. The basic estimation system provides us in training data set. The basic estimation system is implemented using LS-SVMlab1.8 Toolbox [29] in Mat lab environment.

A. Results of the Basic Estimation System

When an utterance of an unseen speaker is applied to a basic estimator, its output is the average weight of speakers in training data set. The basic estimation system provides us a chance level accuracy. The results of speaker weight estimation using a basic estimator are reported in the first row of Table 1. Besides providing a reference level for speaker weight estimation systems, the basic estimator highlights a limitation of using mean-absolute-error (MAE) as a performance metric for weight estimation problem. The MAE is limited in some respects, specially, in the case of a test set with a skewed distribution which is the case in this task. When a test data set with a skewed distribution is applied to a basic estimator, the MAE might be in an acceptable range, based on the variance of the data. For instance, when the database described in Section 4-A was applied to the basic estimator, the MAE for male and female speakers were 12.93 kg and 9.76 kg, respectively. However, the measured CC for males and females were equal to zero. For this reason, the correlation coefficient is a preferred performance metric in this task, which reflects the performance of the estimators in a more sensible way.

B. Results of the i-vector-based System

In this study, the development set consists of utterances of 20% of speakers in training set. In this approach, as illustrated in Fig. 4, each utterance of train, development and test sets are converted to 400 dimensional i-vectors and 300 dimensional NFA vectors. Then, the obtained i-vectors and NFA vectors of the train set are employed to train the i-vector-based and the NFA-based models, respectively. In the next step, the i-vectors and NFA vectors of the development set are applied to the trained models. The outputs are then concatenated to form 2-dimensional vectors and along with the corresponding weight labels are used to train the fusion model. The fusion model is a single hidden layer feed forward neural network with 5 hidden units.

C. The Results of the Score-level Fusion System

In score-level fusion, in which the outputs of different estimators are fused, we need to allocate a portion of training data for development set to train the fusion model. In this study, the development set consists of utterances of 20% of speakers in training set. In this approach, as illustrated in Fig. 4, each utterance of train, development and test sets are converted to 400 dimensional i-vectors and 300 dimensional NFA vectors. Then, the obtained i-vectors and NFA vectors of the train set are employed to train the i-vector-based and the NFA-based models, respectively. In the next step, the i-vectors and the NFA vectors of development set are applied to the trained models. The outputs are then concatenated to form 2-dimensional vectors and along with the corresponding weight labels are used to train the fusion model. The fusion model is a single hidden layer feed forward neural network with 5 hidden units.

Table 1. The MAE (in kg) and CC of the proposed speaker weight estimation systems, compared to the basic and i-vector-based systems.

<table>
<thead>
<tr>
<th>Speaker Weight Estimation System</th>
<th>MALE</th>
<th>FEMALE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>MAE</td>
</tr>
<tr>
<td>Basic estimator</td>
<td>0</td>
<td>12.93</td>
</tr>
<tr>
<td>i-vector-based system</td>
<td>0.42</td>
<td>12.17</td>
</tr>
<tr>
<td>Score-level fusion subsystems</td>
<td>0.43</td>
<td>11.98</td>
</tr>
<tr>
<td>Feature-level fusion-based system</td>
<td>0.56</td>
<td>11.16</td>
</tr>
</tbody>
</table>

Fig. 4. Block diagram of the score-level fusion speaker weight estimation system (U. M. stands for utterance modeling).
Logistic and linear activation functions are considered for the hidden and output neurons, respectively. The network is trained using the one step secant back-propagation algorithm [30] which is implemented using Neural Network Toolbox [31] in Mat lab environment.

The results of the proposed score-level fusion system for speaker weight estimation are presented in the third row of Table 1. Comparing to the results of the i-vector-based estimator, it can be concluded that fusion of the outputs of two subsystems can slightly improve the estimation accuracy which indicates that GMM weights carry complementary information to GMM means. The achieved relative improvements in CC by the proposed fusion scheme compared to the i-vector-based estimator for male and female speakers are 2.32% and 6.25%, respectively.

D. Results of the Proposed Approach

To improve the estimation accuracy of the i-vector-based weight estimation, a feature-level fusion of the i-vectors and the NFA vectors is considered in this paper. In the proposed method, the extracted i-vectors and NFA vectors are length normalized and concatenated to form a longer vector. The obtained vector, along with the corresponding weight label is then used to train estimator. The last row of Table 1 contains the results of the proposed weight estimation approach using a fusion of the i-vectors and the NFA vectors. The obtained results indicate that the accuracy of weight estimation increases after feature-level fusion compared to the estimation using the i-vector-based estimator, which again shows that GMM weights provide complementary information to GMM means.

The achieved relative improvements in CC by the proposed feature-level fusion scheme compared to the i-vector-based estimator for male and female speakers are 25% and 38.77%, respectively. Comparing the results of these two fusion schemes reveals that fusion of the i-vector and the NFA frameworks at feature level is more effective in speaker weight estimation. In addition, fusion at feature level eliminates the need for assigning a considerable amount of training data for development set, and speaker weight estimation is performed in one learning phase.

The reported CC for speaker weight estimation based on the formant parameters of the running speech signals uttered by 91 speakers are 0.33 and 0.34 for male and female speakers, respectively [7]. The results obtained from our proposed speaker weight estimation system seem reasonable, considering the fact that the applied testing dataset in this study consists of spontaneous speech signals and the number of speakers in test set is considerably larger than that of in [7]. It can be concluded that automatic speaker weight estimation using a fusion of the i-vector and the NFA frameworks is more efficient compared to the estimation based on the raw acoustic features.

6. Conclusion

In this paper a novel approach for automatic speaker weight estimation from spontaneous telephone speech signals was proposed. In this method, each utterance was modeled using a fusion of the i-vector and the NFA frameworks at feature level. Using this new utterance modeling approach, the available information in both GMM means and GMM weights was utilized. Then, an LSSVR was employed to estimate the weight of a speaker from a given utterance. The proposed method was trained and tested on the telephone conversations of NIST 2008 and 2010 SRE corpora.

Evaluation results over 2339 utterances show that the correlation coefficients between the actual and the estimated weights of male and female speakers after feature-level fusion are 0.56 and 0.49, respectively, which indicate that the fusion of the i-vectors and the NFA vectors at feature level improves the performance of the state-of-the-art i-vector framework.

Utilizing information in Gaussian weights in conjunction with that of in Gaussian means through a fusion of the i-vector and the NFA frameworks resulted in achieving 25% and 38.77% relative improvements in CC compared to the i-vector-based weight estimation system.

It also indicates the effectiveness of the proposed method in automatic speaker weight estimation compared to the estimation based on the raw acoustic features.

References


Amir Hossein Poorjam was born in Esfahan, Iran, in June 1986. He received his M.Sc. degree in Electrical Engineering from University of Leuven (KU Leuven), Belgium, in 2014. In 2015, he visited the Speech and Image Processing Unit (SIPU), at University of Eastern Finland, where he was granted a CIMO fellowship. He is currently pursuing the Ph.D. degree in the Audio Analysis Laboratory at Aalborg University, Denmark. His research interests include speech signal processing, speaker profiling and language recognition.

Mohamad Hasan Bahari is a postdoctoral researcher at STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven. He received his Ph.D. from KU Leuven, Belgium, where he was granted a Marie-Curie fellowship. During winter, spring and fall 2013, he visited the Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology (MIT). His works on Data Analytics and machine learning was awarded the International Speech Communication Association (ISCA) best student paper at INTERSPEECH 2012 and second place of Reinforcement Learning Competition (Octopus event) 2009.
Hugo Van hamme received the Master’s degree in engineering (burgerlijk ingenieur) from VUB in 1987, the M.Sc. degree from Imperial College, London in 1988 and the Ph.D. degree in electrical engineering from Vrije Universiteit Brussel (VUB) in 1992. From 1993 till 2002, he worked for L&H Speech Products and ScanSoft, initially as senior researcher and later as research manager. Since 2002, he is a professor at the department of electrical engineering of KU Leuven. His main research interests are: applications of speech technology in education and speech therapy, computational models for speech recognition and language acquisition and noise robust speech recognition.
A Novel Approach to Speaker Weight Estimation Using a Fusion of the i-vector and NFA Frameworks

Amir Hossein Poorjam, Mohamad Hasan Bahari, and Hugo Van hamme

Abstract. This paper proposes a novel approach for automatic speaker weight estimation from spontaneous telephone speech signals. In this method, each utterance is modeled using the i-vector framework which is based on the factor analysis on Gaussian Mixture Model (GMM) mean supervectors, and the Non-negative Factor Analysis (NFA) framework which is based on a constrained factor analysis on GMM weight supervectors. Then, the available information in both Gaussian means and Gaussian weights is exploited through a feature-level fusion of the i-vectors and the NFA vectors. Finally, a least-squares support vector regression is employed to estimate the weight of speakers from the given utterances.

The proposed approach is evaluated on spontaneous telephone speech signals of National Institute of Standards and Technology 2008 and 2010 Speaker Recognition Evaluation corpora. To investigate the effectiveness of the proposed approach, this method is compared to the i-vector-based speaker weight estimation and an alternative fusion scheme, namely the score-level fusion. Experimental results over 2339 utterances show that the correlation coefficients between the actual and the estimated weights of female and male speakers are 0.49 and 0.56, respectively, which indicate the effectiveness of the proposed method in speaker weight estimation.

Keywords: I-vector, least-squares support vector regression, non-negative factor analysis, speaker weight estimation.

1. Introduction

The voice of a speaker conveys information about speaker’s traits and states such as age, gender, body size (weight/height) and emotional state [1], [2]. Estimation of speaker’s weight (which is considered as a long term trait of a speaker and an important parameter in various applications) is an interesting and challenging task in forensic, medical and commercial applications. In forensic scenarios, body size estimation of suspects from their voices helps users to receive services proportional to their physical conditions.

The relation between the size of various components of the sound production system (such as vocal folds and vocal tract) and the body size of a speaker has motivated researchers in the field of speaker recognition to look for features of an acoustic signal that provide cues to the body size of speakers. For instance, authors in [3] found a relationship between formants and the length of the vocal tract, based on the source-filter theory. Thus, since the vocal tract is a part of speaker’s body, this feature can be used to estimate the weight of a speaker [4].

However, speaker weight estimation from the voice patterns is challenging. For instance, mean fundamental frequency ($f_0$) of voice is reported as a feature which has a (negative) correlation with body size. That is, females and children have higher $f_0$ while in males (who are taller and heavier), this value is lower [5]. However, when the relation of the fundamental frequency ($f_0$) and weight was investigated within male and female speakers, no correlation was found between $f_0$ and the weight of adult humans [6], [7]. The lowest fundamental frequency of voice ($F_{f0min}$) is another feature which is determined by the mass and length of the vocal folds [6]. By investigating this feature, researchers have found no correlation between $F_{f0min}$ and weight in adult human speakers [6], [7]. Fitch has found formant dispersion (the averaged difference between adjacent pair of formant frequencies) a reliable feature which has a correlation with both vocal tract length and body size in macaques [8]. However, a weak relation between formant parameters and weight of human adults is reported in study conducted by Gonzalez [9]. This week correlation may be due to the fact that the vocal folds in humans at puberty grow independent of the rest of the head and body. This issue is more evident in the males than the females [10], [11]. Gonzalez studied the correlation between formant frequencies and weight in human adults [9]. He calculated the formant parameters by means of a long-term average analysis of running speech signals uttered by 91 speakers. In this experiment, the Pearson correlation coefficients between formants and weights for male and female speakers were reported to be 0.33 and 0.34, respectively [9]. In research conducted by Van Dommelen and Moxness [12], the ability to judge the weight of speakers from their speech samples was investigated. They reported a significant correlation between estimated weight (judged by listeners) and actual weight of only male speakers. In addition, they performed a regression analysis involving several acoustic features such as $f_0$, formant frequencies, energy below 1 kHz, and speech rate. The results showed that the speech rate was the only parameter...
which had a significant correlation with male speaker’s weight. They concluded that speech rate of male speakers is a reliable predictor for weight estimation.

Modeling speech utterances with Gaussian mixture model (GMM) mean supervectors is demonstrated to be an effective approach to speaker recognition [13]. However, GMM mean supervectors are high dimensional vectors, and obtaining a reliable model is difficult when limited data are available. Recently, utterance modeling using the i-vector framework [14] has considerably increased the accuracy of the classification and regression problems in the field of speaker characterization [15], [16], speaker verification [14] and language recognition [17], [18]. The i-vector, which is based on the factor analysis on GMM mean supervectors, represents an utterance in a compact and a low-dimensional feature vector. In addition, various studies show that although GMM weights convey less information than GMM means, they provide complementary information to GMM means [19]–[23]. A Non-negative Factor Analysis (NFA) framework [21], which is based on a constrained factor analysis for GMM weights, has been recently introduced and yields a new low-dimensional utterance representation. In [24], we successfully applied a score-level fusion of the i-vector and the NFA frameworks to simultaneously estimate various characteristics of speakers from speech signals. We showed that utilizing information in both GMM means and GMM weights through a score-level fusion of the i-vector and the NFA frameworks is an effective approach to improve the estimation accuracy. However, the fusion at score level requires a large development data set to train the fusion model, which results in decreasing the number of available training data.

In this study, a new speech-based method for automatic weight estimation is proposed in which instead of using raw acoustic features, each utterance is modeled using a fusion of the i-vector and the NFA frameworks at feature level. In this new utterance modeling approach, in addition to exploiting the available information in GMM means and GMM weights, the need for assigning a considerable amount of training data for development set is eliminated and speaker weight estimation is performed in one learning phase. To perform function approximation, a least-squares support vector regression (LSSVR) is utilized in this paper. For the comparison purpose, the proposed method is compared to the i-vector-based speaker weight estimation and speaker weight estimation using an alternative fusion scheme, namely the score-level fusion. The proposed approach is evaluated on spontaneous telephone speech signals of the NIST 2008 and 2010 SRE corpora. Experimental results confirm the effectiveness of the proposed approach in automatic speaker weight estimation.

The rest of the paper is organized as follows. In Section 2 the problem of automatic weight estimation is formulated and different baseline systems for speaker weight estimation are described. The proposed approach is elaborated in Section 3. Section 4 explains the experimental setup. The evaluation results are presented and discussed in Section 5. The paper ends with conclusions in Section 6.

2. Automatic Weight Estimation from Speech Signals

In this section, the problem of automatic weight estimation is formulated and different baseline approaches are described.

A. Problem Formulation

In the speaker weight estimation problem, we are given a set of training data \( D = \{O_i, y_i\}_{i=1}^{n} \), where \( O_i \) denotes the \( i \)-th utterance and \( y_i \) denotes the corresponding weight.

The goal is to approximate a function \( g \), such that for an utterance of an unseen speaker, \( O_{\text{test}} \), the estimated weight, \( \hat{y} = g(O_{\text{test}}) \), approximates the actual weight as good as possible.

B. Baseline Approaches

For the comparison purpose, the proposed method is compared to three baseline approaches, namely the basic estimator, the i-vector-based speaker weight estimation [23] and speaker weight estimation using score-level fusion of the i-vector and the NFA frameworks [24].

1) Basic Estimation System:

The output of a basic estimator is the average weight of speakers in training data set. The basic estimation system provides us a chance level accuracy.

2) The i-vector-based System:

In the i-vector-based weight estimation system, each utterance is mapped onto a 400 dimensional vector using the i-vector framework. Then, the extracted i-vectors along with their corresponding weight labels are used to train estimator. This method is considered as the baseline system in this paper.

3) The Score-Level Fusion System:

The fusion of the i-vector and the NFA frameworks is an effective approach to exploit the available information in both GMM means and GMM weights and consequently to enhance the estimation accuracy. In this method, each utterance is converted to an i-vector and an NFA vector. Then, the obtained vectors of the train set are employed to train the i-vector-based and the NFA-based models. In the next step, the i-vectors and the NFA vectors of development set are applied to the trained models. The outputs are then concatenated to form 2-dimensional vectors and along with the corresponding weight labels are used to train the fusion model. This system, which is labeled as the score-level fusion system in this paper, is presented to investigate the effect of different fusion schemes in speaker weight estimation.

3. System Description

A. Utterance Modeling

By fitting a GMM to acoustic features extracted from each speech signal, a variable-duration speech signal is converted into a fixed-dimensional vector which is suitable for regression algorithms. The parameters of the obtained GMM characterize the corresponding utterance. Due to limited data, we are not able to accurately fit a separate GMM for a short utterance, especially in the case of GMMs
with a high number of Gaussian components. Thus, for adapting a universal background model (UBM) to characteristics of utterances in training and testing databases, parametric utterance adaptation techniques are applied. In this paper, the i-vector and the NFA frameworks are applied to adapt UBM means and weights, respectively.

1) Universal Background Model and Adaptation:

Consider a UBM with the following likelihood function of data \( O = \{ o_1, \ldots, o_r \} \):

\[
p(o_t | \gamma) = \sum_{c} \pi_c p(o_t | \mu_c, \Sigma_c)
\]

\( \gamma = [\pi_c, \mu_c, \Sigma_c], c = 1, \ldots, C \) (1)

where \( o_t \) is the acoustic vector at time \( t \), \( \pi_c \) is the mixture weight for the \( c^{th} \) mixture component, \( p(o_t | \mu_c, \Sigma_c) \) is a Gaussian probability density function with mean \( \mu_c \) and covariance matrix \( \Sigma_c \), and \( C \) is the total number of Gaussian components in the mixture. The parameters of the UBM – \( \gamma \) – are estimated on a large amount of training data.

2) The i-vector Framework:

One effective method for speaker weight estimation involves adapting UBM means to the speech characteristics of the utterance. Then, the adapted GMM means are extracted and concatenated to form Gaussian mean supervectors. However, since Gaussian components of the UBM model are adapted independent of each other, some components are not updated in the case of limited training samples [25]. This problem can be alleviated by linking the Gaussian components together using the Joint Factor Analysis (JFA) framework [26].

In the JFA framework, each utterance is represented by a supervector \( M \) which is a speaker- and channel-dependent vector of dimension (C,F), where \( C \) is the total number of the mixture components in a feature space of dimension \( F \). In the JFA framework, it is assumed that \( M \) can be decomposed into two supervectors:

\[
M = s + e
\]

(2)

where \( s = u + Vq + Dr \) is a speaker-dependent supervector and \( e = Up \) is a channel-dependent supervector. \( s \) and \( e \) are independent and possess normal distributions. \( u \) is the speaker- and channel-independent supervector, \( V \) defines a lower dimensional speaker subspace, \( U \) is a lower dimensional channel subspace, and \( D \) defines a speaker subspace. \( q \) and \( r \) are factors in speaker subspace, and \( p \) is a channel-dependent factor in channel subspace. The vectors \( p, q \) and \( r \) are random variables with standard normal distributions \( N(0, I) \) which are jointly estimated.

In the JFA framework, the channel factor contains some information about speakers, which can be utilized in speaker identification. This fact resulted in proposing a new utterance modeling approach, referred to as the i-vector framework or the total variability modeling [27]. This method comprises both speaker variability and channel variability. Channel compensation procedures such as within-class covariance normalization can be further applied to compensate the residual channel effects in the speaker factor space [28].

The i-vector framework assumes that each utterance possesses a speaker- and channel-dependent GMM supervector which its mean, \( M \), can be decomposed as

\[
M = u + Tv
\]

(3)

where \( u \) is the mean supervector of the UBM, and \( T \) spans a low-dimensional subspace (400 dimensions in this work). In the i-vector framework, \( T \) and \( v \) are estimated using the Expectation-Maximization (EM) algorithm. In the E-step, \( T \) is supposed to be known, and \( v \) is updated. In the M-step, \( v \) is assumed to be known, and \( T \) is updated. The subspace vector \( v \) is treated as a hidden variable with the standard normal prior and the i-vector is its maximum-a-posteriori (MAP) point estimate which is calculated by maximization of the following auxiliary function over \( v \):

\[
\Psi(\gamma, v) = \sum_{c} \theta_{c,t} \log p(o_t | [\mu_c + T_v, v], \Sigma_c)N(v)
\]

(4)

where \( N(v) \) denotes the standard normal distribution of \( v \), \( T_c \) are the rows of the subspace matrix \( T \), which correspond to the \( c^{th} \) Gaussian mean, and \( \theta_{c,t} \) is the occupation count for the \( c^{th} \) mixture component and \( t^{th} \) frame. The occupation count is calculated as follows:

\[
\theta_{c,t} = \frac{\pi_c p(o_t | \mu_c, \Sigma_c)}{\sum_c \pi_c p(o_t | \mu_c, \Sigma_c)}
\]

(5)

In the E-step, the posterior distribution of \( v \) is Gaussian with the following mean \( \mu_v \) and covariance matrices \( \Sigma_v \) [29]:

\[
\mu_v = \left[ I + \sum_c \theta_{c,t} T_c^T \Sigma_c T_c \right]^{-1} T_c^T \Sigma_c^{-1} \sum_c \theta_{c,t} (o_t - \mu_c)
\]

(6)

\[
\Sigma_v = \frac{1}{\sum_c \theta_{c,t}} \left[ I + \sum_c \theta_{c,t} T_c^T \Sigma_c T_c \right]^{-1}
\]

(7)

where \( I \) denotes an identity matrix of appropriate size, \( u_c \) and \( \Sigma_c \) are adapted mean and covariance of the \( c^{th} \) Gaussian, which are updated during each EM iteration starting from UBM parameters, and \( T \) represents the transpose operator.

In the M-step, the subspace matrix \( T \) is estimated via maximization of the following auxiliary function over \( T \):

\[
\bar{\Psi}(\gamma, T) = \sum_{c} \sum_{t} \theta_{c,t} \log p(o_t | [\mu_c + T_v, v], \Sigma_c)
\]

(8)

An efficient procedure for training \( T \) and for MAP adaptation of the i-vectors can be found in [29]. In the total variability modeling approach, the i-vector is the low-dimensional representation of an audio recording that can be used for classification and estimation purposes.

3) The Non-negative Factor Analysis (NFA) Framework:

The NFA is a new framework for adaptation and decomposition of GMM weights based on a constrained factor analysis [21]. The basic assumption of this method is that for a given utterance, the adapted GMM weight supervector can be decomposed as follows:

\[
w = \pi + Lr.
\]

(9)
where $\pi$ is the UBM weight supervector (2048 dimensional vector in this study). $L$ is a matrix of dimension $C \times \rho$ spanning a low-dimensional subspace (300 dimensions in this work). $r$ is a low-dimensional subspace vector obtained through a constrained maximum likelihood estimation criterion.

In this framework, the adapted weights are obtained by maximizing the following objective function over $w_c$.

$$
\Psi(y, r) = \sum_{c=1}^{C} \sum_{t=1}^{C} \log \pi \cdot p(O_t | \mu, \Sigma) \tag{10}
$$

Substituting $w_c$ by $(\pi + L \cdot r)$ in the Eq. 10, and given an utterance $O$, a maximum likelihood estimation of $r$ is obtained by solving the following constrained optimization problem:

$$
\max_{r} \Psi(y, r) = \max_{r} \left[ (\pi + Lr)^{\top} \log (\pi + Lr) \right] \tag{11}
$$

Subject to

$$
|Lr + \pi| > 0
$$

where $L$ is a row vector of dimension $C$ with all elements equal to one, and $\log(O) = \sum [\theta_{r_1, \ldots, \theta_{r_C}]^T$.

In this framework, neither the subspace matrix $L$ nor the subspace vector $r$ is constrained to be non-negative. However, unlike the i-vector framework, the applied factor analysis for estimating the subspace matrix $L$ and the subspace vector $r$ is constrained such that the adapted GMM weights are non-negative and sum up to one. The procedure of calculating $L$ and $r$ involves a two-stage algorithm similar to EM and can be found in [21]. The subspace matrix $L$ is estimated over a large training dataset. It is then used to extract a subspace vector $r$ for each utterance in train and test datasets.

This new low-dimensional utterance representation approach was successfully applied to speaker characterization [23], [24] and language/dialect recognition [21] tasks.

4) Feature-Level Fusion of the i-vectors and NFA Vectors:

Previous studies show that although GMM weight supervectors contain less information than GMM means, they provide complementary information to GMM means [22]. Feature-level fusion and score-level fusion are considered as effective approaches to exploit available information in both GMM means and weights [22], [24]. Score-level fusion, in which the outputs of different estimators are fused, requires a development data set to train the fusion model, which results in decreasing the number of training data. However, fusion at feature level, in which various features are normalized and concatenated, eliminates the need for assigning a considerable amount of available training data for development set, and estimation can be performed in one learning phase.

In this paper, a feature-level fusion of the i-vectors and the NFA vectors is considered to improve the estimation accuracy. As illustrated in Fig. 1, the extracted i-vectors and the NFA vectors are length normalized by having mapped onto a low-dimensional space using linear discriminant analysis (LDA) [30]. Then, the obtained low-dimensional vectors are concatenated to form a longer vector.

B. Function Approximation Using LSSVR

Support vector regression (SVR) is a function approximation approach developed as a regression version of the widely known Support Vector Machines (SVM) classifier. Using nonlinear transformations, SVMs map the input data onto a higher dimensional space in which a linear solution can be calculated. They also keep a subset of the samples which are the most relevant data for the solution and discard the rest. This makes the solution as sparse as possible. While SVMs perform the classification task by determining the maximum margin separation hyperplane between classes, SVR carries out the regression task by finding the optimal regression hyperplane in which most of training samples lie within an $\varepsilon$-margin around this hyperplane [31].

In this study, we use the least squares version of support vector regression. While an SVR solves a quadratic programming with linear inequality constraints, which results in high algorithmic complexity and memory requirement, an LSSVR involves solving a set of linear equations by considering equality constraints instead of inequalities for classical SVR [31], which speeds up the calculations.

In a regression problem, we are given a training dataset $D = \{(x_1, y_1), \ldots, (x_k, y_k)\}$, where $x_k$ and $y_k$ denote a vector of observed features of the $k$th training item and its corresponding output, respectively. The goal is to determine a function $h(o)$ such that the outputs are predicted accurately. In primal form of LSSVR, $h(o)$ is considered as

$$
h(o) = \beta^{\top} \phi(o) + c \tag{12}
$$

A least squares loss function is applied instead of Vapnik’s $\varepsilon$-insensitive loss function in LSSVR to simplify the formulation by minimizing

$$
\frac{1}{2} \|w\|^2 + \frac{1}{2} \sum_{i=1}^{N} e_i^2 \tag{13}
$$

subject to

$$
y_i = \beta^{\top} \phi(o_i) + c + e_i \tag{14}
$$

where $\beta$ is an error cost factor and $e_i \in \mathbb{R}$ are error variables.

This optimization problem can be solved more efficiently for high dimensional data by using the Lagrangian variables $\nu$ and minimizing the following dual cost function [31].

$$
\Omega(\vartheta, c, e, \nu) = \frac{1}{2} \|\vartheta\|^2 + \frac{1}{2} \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \nu_i \left[ \beta^{\top} \phi(o_i) + c + e_i - y_i \right] \tag{15}
$$

![Fig. 1. Block diagram of the utterance modeling in feature-level fusion.](image-url)
This minimization problem can be solved directly by taking the partial derivative of $\Omega$ with respect to $\theta$, $c$, $e$ and $\nu$ and setting the results to zero. This results in solving a linear system of equations. Inserting the obtained results to (12) leads to the regression function

$$h(o) = \sum_{n=1}^{N} v_n \phi(o_n, o) + c = \sum_{n=1}^{N} v_n K(o_n, o) + c$$

(16)

where $K(o_n, o)$ denotes the kernel function and $c$ and $\nu$ are the solution to optimization problem (15).

One drawback of the applied simplification in LSSVR formulation is the loss of sparseness. Therefore, all samples contribute to the model, and consequently, the model often becomes unnecessarily large.

C. Training and Testing

The block diagram of the proposed weight estimation approach is shown in Fig. 2. During the training phase, each utterance in the training data set is mapped to a high dimensional vector using the feature-level fusion utterance modeling described in Section 3.A.4. Then, the obtained vectors along with their corresponding weight labels are used to train an estimator for approximating function $g$. During the testing phase, the same utterance modeling approach applied in training phase is used to extract a high dimensional vector from a test utterance. Then, the estimated weight is obtained using the trained estimator.

4. Experimental Setup

A. Database

The National Institute for Standard and Technology (NIST) has held annual or biannual speaker recognition evaluations (SRE) for the past two decades. With each SRE, a large corpus of telephone (and more recently microphone) conversations is released. Conversations typically last 5 minutes and originate from a large number of speakers for whom additional meta-data is recorded.

The NIST databases were chosen for this work due to the large number of speakers and because the total variability subspace requires a considerable amount of development data for training. The development data set used to train the total variability subspace and UBM includes over 30,000 speech recordings and was sourced from the NIST 2004-2006 SRE databases, LDC releases of Switchboard 2 phase III and Switchboard Cellular (parts 1 and 2).

For the purpose of automatic speaker weight estimation, telephone recordings from the common protocols of the recent NIST 2008 and 2010 SRE databases are pooled together to create a dataset of 8241 utterances uttered by 1333 speakers. Then, it is divided to two disjoint parts such that 80% and 20% of all speakers are used for training and testing sets, respectively. Thus, all 8241 utterances, 5902 utterances are considered for training set and 2339 utterances are considered for testing set. Fig. 3 shows the weight histograms of training and testing datasets for male and female speakers.

B. Performance Metric

In order to evaluate the effectiveness of the proposed method, the mean-absolute-error (MAE) of the estimated weight and the Pearson correlation coefficient (CC) between the actual and estimated weights are used. MAE is defined as:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

(17)

where $\hat{y}_i$ is the $i^{th}$ estimated weight, $y_i$ is the $i^{th}$ actual weight, and $N$ is the total number of test samples.

The Pearson correlation coefficient is computed as:

$$CC = \frac{1}{N-1} \sum_{i=1}^{N} \left( \frac{\hat{y}_i - \mu_\hat{y}}{\sigma_\hat{y}} \right) \left( \frac{y_i - \mu_y}{\sigma_y} \right)$$

(18)

where $\mu_\hat{y}$ and $\sigma_\hat{y}$ denote the mean and standard deviation of the actual speakers’ weight respectively, and $\mu_y$ and $\sigma_y$ are respectively the mean and standard deviation of the estimated weights.

5. Results and Discussion

In this section, the evaluation results of the baseline systems as well as the proposed speaker weight estimation approach are presented. The acoustic feature vector is a 60-dimensional vector consists of 20 Mel-Frequency Cepstrum Coefficients (MFCCs) including energy appended with their first and second order derivatives. This type of feature is very common in the i-vector-based speaker recognition systems. Wiener filtering, feature warping [32]
and voice activity detection [33] have also been considered in the front-end processing to obtain more reliable features.

In this study, an LSSVR with a linear kernel has been employed to perform weight estimation, which is implemented using LS-SVMLab1.8 Toolbox [34] in Matlab environment.

A. Results of the Basic Estimation System

When an utterance of an unseen speaker is applied to a basic estimator, its output is the average weight of speakers in training data set. The basic estimation system provides us a chance level accuracy. The results of speaker weight estimation using a basic estimator are reported in the first row of Table I. Besides providing a reference level for speaker weight estimation systems, the basic estimator highlights a limitation of using mean-absolute-error (MAE) as a performance metric for weight estimation problem. The MAE is limited in some respects, specially, in the case of a test set with a skewed distribution which is the case in this task. When a test data set with a skewed distribution is applied to a basic estimator, the MAE might be in an acceptable range, based on the variance of the data. For instance, when the database described in Section IV-A was applied to the basic estimator, the MAE for male and female speakers were 12.93 kg and 9.76 kg, respectively. However, the measured CC for males and females were equal to zero. For this reason, the correlation coefficient is a preferred performance metric in this task, which reflects the performance of the estimators in a more sensible way.

B. Results of the i-vector-Based System

In the i-vector-based system, each utterance in training set is mapped onto a 400 dimensional vector using the i-vector framework. Then, the extracted i-vectors along with their corresponding weight labels are used to train estimator. The results of employing an LSSVR as an estimator, and using the i-vector framework for utterance modeling are presented in the second row of Table I. Comparing to the results of the basic estimator, the MAE might be in an acceptable range, based on the variance of the data. For instance, when the database described in Section IV-A was applied to the basic estimator, the MAE for male and female speakers were 12.93 kg and 9.76 kg, respectively. However, the measured CC for males and females were equal to zero. For this reason, the correlation coefficient is a preferred performance metric in this task, which reflects the performance of the estimators in a more sensible way.

C. The Results of the Score-Level Fusion System

In score-level fusion, in which the outputs of different estimators are fused, we need to allocate a portion of training data for development set to train the fusion model. In this study, the development set consists of utterances of 20% of speakers in training set. In this approach, as illustrated in Fig. 4, each utterance of train, development and test sets are converted to 400 dimensional i-vectors and 300 dimensional NFA vectors. Then, the obtained i-vectors and NFA vectors of the train set are employed to train the i-vector-based and the NFA-based models, respectively. In the next step, the i-vectors and the NFA vectors of development set are applied to the trained models. The outputs are then concatenated to form 2-dimensional vectors and along with the corresponding weight labels are used to train the fusion model. The fusion model is a single hidden layer feedforward neural network with 5 hidden units. Logistic and linear activation functions are considered for the hidden and output neurons, respectively. The network is trained using the one step secant back-propagation algorithm [35] which is implemented using Neural Network Toolbox [36] in Matlab environment.

The results of the proposed score-level fusion system for speaker weight estimation are presented in the third row of Table I. Comparing to the results of the i-vector-based estimator, it can be concluded that fusion of the outputs of two subsystems can slightly improve the estimation accuracy which indicates that GMM weights carry complementary information to GMM means. The achieved relative improvements in CC by the proposed fusion scheme compared to the i-vector-based estimator for male and female speakers are 2.32% and 6.25%, respectively.

D. Results of the Proposed Approach

To improve the estimation accuracy of the i-vector-based weight estimation, a feature-level fusion of the i-vectors and the NFA vectors is considered in this paper. In the proposed method, the extracted i-vectors and NFA vectors are length normalized and concatenated to form a longer vector. The obtained vector, along with the corresponding weight label is then used to train estimator. The last row of Table 1 contains the results of the proposed weight estimation approach using a fusion of the i-vectors and the NFA vectors. The obtained results indicate that the accuracy of weight estimation increases after feature-level fusion compared to the estimation using the i-vector-based estimator, which again shows that GMM weights provide complementary information to GMM means.

![Fig. 4. Block diagram of the score-level fusion speaker weight estimation system (U. M. stands for utterance modeling).](image)
The achieved relative improvements in CC by the proposed feature-level fusion scheme compared to the i-vector-based estimator for male and female speakers are 25% and 38.77%, respectively. Comparing the results of these two fusion schemes reveals that fusion of the i-vector and the NFA frameworks at feature level is more effective in speaker weight estimation. In addition, fusion at feature level eliminates the need for assigning a considerable amount of training data for development set, and speaker weight estimation is performed in one learning phase.

The reported CC for speaker weight estimation based on the formant parameters of the running speech signals uttered by 91 speakers are 0.33 and 0.34 for male and female speakers, respectively [9]. The results obtained from our proposed speaker weight estimation system seem reasonable, considering the fact that the applied testing dataset in this study consists of spontaneous speech signals and the number of speakers in test set is considerably larger than that of in [9]. It can be concluded that automatic speaker weight estimation using a fusion of the i-vector and the NFA frameworks is more efficient compared to the estimation based on the raw acoustic features.

6. Conclusion

In this paper a novel approach for automatic speaker weight estimation from spontaneous telephone speech signals was proposed. In this method, each utterance was modeled using a fusion of the i-vector and the NFA frameworks at feature level. Using this new utterance modeling approach, the available information in both GMM means and GMM weights was utilized. Then, an LSSVR was employed to estimate the weight of a speaker from a given utterance. The proposed method was trained and tested on the telephone conversations of NIST 2008 and 2010 SRE corpora. Evaluation results over 2339 utterances show that the correlation coefficients between the actual and the estimated weights of male and female speakers after feature-level fusion are 0.56 and 0.49, respectively, which indicate that the fusion of the i-vectors and the NFA vectors at feature level improves the performance of the state-of-the-art i-vector framework. Utilizing information in Gaussian weights in conjunction with that of in Gaussian means through a fusion of the i-vector and the NFA frameworks resulted in achieving 25% and 38.77% relative improvements in CC compared to the i-vector-based weight estimation system. It also indicates the effectiveness of the proposed method in automatic speaker weight estimation compared to the estimation based on the raw acoustic features.

References


[17] A. H. Poorjam, R. Saeidi, T. Kinnunen, and V. Hautam, “Incorporating uncertainty as a quality measure in i-


Amir Hossein Poorjam was born in Esfahan, Iran, in June 1986. He received his M.Sc. degree in Electrical Engineering from University of Leuven (KU Leuven), Belgium, in 2014. In 2015, he visited the Speech and Image Processing Unit (SIPU), at University of Eastern Finland, where he was granted a CIMO fellowship. He is currently pursuing the PhD degree in the Audio Analysis Laboratory at Aalborg University, Denmark. His research interests include speech signal processing, speaker profiling and language recognition.

Mohamad Hasan Bahari is a postdoctoral researcher at STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven. He received his PhD from KU Leuven, Belgium, where he was granted a Marie-Curie fellowship. During winter, spring and fall 2013, he visited the Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology (MIT). His works on Data Analytics and machine learning was awarded the International Speech Communication Association (ISCA) best student paper at INTERSPEECH 2012 and second place of Reinforcement Learning Competition (Octopus event) 2009.

Hugo Van hamme received the Master’s degree in engineering (burgerlijk ingenieur) from VUB in 1987, the M.Sc. degree from Imperial College, London in 1988 and the Ph.D. degree in electrical engineering from Vrije Universiteit Brussel (VUB) in 1992. From 1993 till 2002, he worked for L&H Speech Products and ScanSoft, initially as senior researcher and later as research manager.
Since 2002, he is a professor at the department of electrical engineering of KU Leuven. His main research interests are: applications of speech technology in education and speech therapy, computational models for speech recognition and language acquisition and noise robust speech recognition.
Aims and Scope

The Journal of Electrical Systems and Signals (JESS) is devoted to report original research in all aspects of electrical systems and signals. This includes, but not limited to the following areas:

- Biomedical Systems and Signals
- Electronic Systems and Signals
- Communication Systems and Signals
- Control Systems and Signals
- Power Engineering Systems and Signals

Three types of papers are considered for publication in JESS: full papers, technical notes, and review papers. A full paper is a complete report of an original work and normally is between five and ten pages. A technical note is an early report of a new finding and it is normally less than four pages. A review article is an informative paper and includes all aspects of a research topic, such as history, works done before, and the going on activities. This type of paper is normally more than ten pages and is invited by the journal. However, review papers submitted by distinguished researchers in their expertise are also accepted.
**Information for Contributors**

The editorial board of JESS welcomes papers that report original research within the scope of the journal. The submitted papers should not be published before or currently be under the review process elsewhere. Extended versions of conference papers are exempted and will be reviewed, provided the copy right is honored.

The papers should be prepared according to the instructions provided below and submitted using the Web site of the journal.

The manuscript should have two columns of 8.5 cm width in each page and be prepared by the Microsoft Word software using the Times New Roman fonts. The following points should be considered in preparing the manuscript.

1. The title of the paper should be concise, but expressive.

2. The manuscript must contain an abstract of up to 200 words followed by up to 8 keywords.

3. The main body of the paper should be divided into sections and subsections and also be numbered.

4. The references should be numbered according to their appearance in the text, and the information about the references should be provided according to the IEEE standards.

5. The figures and tables should be prepared in the final size, such that no enlargement or reduction is needed. The figures should have captions below and the tables should have headings above.

6. The affiliations of the authors and the email address of the corresponding author should be provided at the footnote of the first page.

7. To provide more information about the authors, it is requested that a brief professional biography (not exceeding 150 words) and a photograph of each author be included at the end of the paper.