A Diffusion Least-Mean Square Algorithm for Distributed Estimation over Sensor Networks

Amir Rastegarnia, Mohammad Ali Tinati and Azam Khalili

Abstract—In this paper we consider the issue of distributed adaptive estimation over sensor networks. To deal with more realistic scenario, different variance for observation noise is assumed for sensors in the network. To solve the problem of different variance of observation noise, the proposed method is divided into two phases: I) Estimating each sensor's observation noise variance and II) using the estimated variances to obtain the desired parameter. Our proposed algorithm is based on a diffusion least mean square (LMS) implementation with linear combiner model. In the proposed algorithm, the step-size parameter the coefficients of linear combiner are adjusted according to estimated observation noise variances. As the simulation results show, the proposed algorithm considerably improves the diffusion LMS algorithm given in literature.

Keywords—adaptive filter, distributed estimation, sensor network, diffusion.

I. INTRODUCTION

Sensor networks have broad range of applications including information gathering, precision agriculture, environment surveillance, target localization and etc. [1]. In all the supposed applications for sensor networks, distribution of the nodes yields spatial diversity, which should be exploited alongside the temporal dimension in order to enhance the robustness of the processing tasks [2]. The ability to detect events of interest is essential to the success of emerging sensor network technology. Detection often serves as the initial goal of a sensing system [1]. Recently, distributed adaptive estimation algorithms that enable a network of nodes to function as an adaptive entity are proposed [3]-[11].

In [3]-[6] distributed adaptive algorithm using incremental optimization techniques such as incremental distributed LMS, (IDLMS) [3-5] and distributed recursive least mean square (dRLS) [6] are developed. The resulting algorithms are distributed, cooperative, and able to respond in real time to changes in the environment. In these algorithms, each node is allowed to communicate with its immediate neighbor in order to exploit the spatial dimension while limiting the communications burden at the same time. In [7-9] diffusion implementation for distributed adaptive estimation algorithms are developed. In these algorithms each node can communicate with all its neighbors as dictated by the network topologyare. In comparison with *incremental* based algorithms, *diffusion* based methods need more communication resources and have

better estimation performance. Both diffusion LMS and diffusion RLS algorithm are introduced in the literature. Their formulation and performance can be found in [10] and [11].

In the existing distributed adaptive estimation algorithms, either observation noise with equal variance is assumed for all the nodes in the network or same strategy is used for different variance condition. In practice, this assumption fails due to the many physical problems. However, considering different observation noise is a better assumption for sensor networks and is more close to real scenario. In this paper, the issue of distributed adaptive estimation over sensor networks is considered with different variance of observation noise constraint. To deal with this constraint, we first obtain an estimate of each sensor's observation noise variance using the diffusion LMS algorithm. In the next step, we propose a new diffusion LMS based algorithm in which the step-size parameter and the coefficients of linear combiner are adjusted according to estimated observation noise variances.

II. ESTIMATION PROBLEM AND THE ADAPTIVE DISTRIBUTED SOLUTION

A. Notation

A list of the symbols used through the paper, for ease of reference, are shown in Table I.

TABLE I
LIST OF THE MAIN USED SYMBOLS

Symbol	Description
w_i	Weight vector estimate at iteration i
u_i	Regressor vector at iteration i
$e(i)$	Output estimation at iteration i
$d(i)$	Value of a *scalar* variable d at iteration i
u_i	Value of a *vector* variable u at iteration i

B. problem Statement

We are interested in estimating an unknown vector w^o from multiple spatially independent but possibly time-correlated measurements collected at N nodes in a network [3] (See Fig. 1). Each node k has access to time-realizations $\{d_k(i), u_{k,i}\}$ of zero-mean spatial data $\{d_k, u_k\}$ where each d_k is a scalar measurement and each u_k is a $1 \times M$ row regression vector.

Amir Rastegarnia is with the Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran, email: a_rastegar@ieee.org

Azam Khalili is with the Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran, email: a-khalili@tabrizu.a.cir

Mohammad Ali Tinati is Professor and Chairman of Electrical Engineering with the University of the of Tabriz, Tabriz, Iran, email: tinati@tabrizu.ac.ir
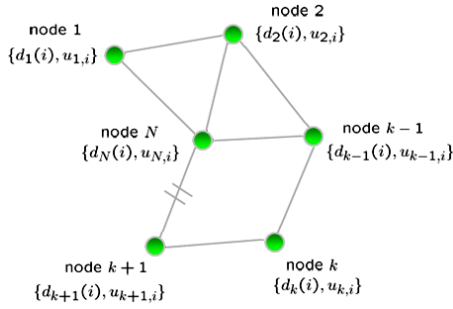
Fig. 1. A distributed network with N active nodes accessing space-time data.

We assume that the unknown vector relates to the as:

$$d_k(i) = u_{k,i} w^o + v_k(i), \qquad (1)$$

where is some white noise sequence (also known as observation noise) with variance , and is independent of . By collecting regression and measurement data into global matrices results:

$$U \triangleq col\{u_1, u_2, ..., u_N\} \qquad (2)$$

$$d \triangleq col\{d_1, d_2, ..., d_N\} \qquad (3)$$

where the notation $col\{\cdot\}$ denotes a column vector (or matrix) with the specified entries stacked on top of each other. The objective is to estimate the vector that solves

$$\min_w J(w) \ where \ J(w) = E \|d - Uw\|^2 \qquad (4)$$

The optimal solution satisfies the normal equations [8]

$$R_{du} = R_u w^o \qquad (5)$$

where

$$R_{du} = E\{U^* d\} = \sum_{k=1}^{N} R_{du,k}, \qquad (6)$$

$$R_{uu} = E\{U^* U\} = \sum_{k=1}^{N} R_{u,k}. \qquad (7)$$

where in (6), the symbol * denotes the Hermitian transform. Note that in order to use (5) to compute w^o each node must have access to the global statistical information $\{R_u, R_{du}\}$ which in turn requires more communications between nodes and computational resources.

C. Incremental LMS solution

The standard gradient-descent implementation to solve the normal equation (5) is as

$$w_i = w_{i-1} + \mu [\nabla J(w_{i-1})]^*, \qquad (8)$$

where μ is a suitably chosen step-size parameter, w_i is an estimate for desired parameter (i.e. w_o) in ith iteration of $\nabla J(\cdot)$ and denotes the gradient vector of $J(w)$ evaluated at w_{i-1}. If μ is sufficiently small then $w_i \to w^o$ as $i \to \infty$

[3-6]. In order to obtain a distributed version of (8), first the cost function $J(w)$ is decomposed as

$$J(w) = \sum_{k=1}^{N} J_k(w), \qquad (9)$$

where

$$J_k \triangleq E\left\{|d_k - U_k w|^2\right\}. \qquad (10)$$

Using (9) and (10) the standard gradient-descent implementation of (8) can be rewritten as [3-6]:

$$w_i = w_{i-1} - \mu \left[\sum_{k=1}^{N} \nabla J_k(w_{i-1})\right]^* \qquad (11)$$

By defining the as the local estimate of the $\psi_k^{(i)}$ at node k and time i, then w_i can be evaluated as [3]

$$\psi_k^{(i)} = \psi_{k-1}^{(i)} - \mu [\nabla J_k(w_{i-1})]^*, \ k = 1, 2, \ldots, N \qquad (12)$$

This scheme still requires all node to share global information w_{i-1}. The fully distributed solution can be achieved by using the local estimate $\psi_k^{(i)}$ at each node k instead of w_{i-1},

$$\psi_k^{(i)} = \psi_{k-1}^{(i)} - \mu \left[\nabla J_k\left(\psi_k^{(i)}\right)\right]^*, \ k = 1, 2, \ldots, N \qquad (13)$$

Now, we need to determine the gradient of J and replace it in (13). To do this, the following approximations are used

$$R_{du,k} \approx d_k(i) u_{k,i}^* \qquad (14)$$

$$R_{u,k} \approx u_{k,i}^* u_{k,i} \qquad (15)$$

The resulting IDLMS algorithm is as follows [5]

$$\begin{cases} \psi_0^{(i)} \leftarrow w_{i-1} \\ \psi_k^{(i)} = \psi_{k-1}^{(i)} - \mu u_{k,i}^* \left[d_k(i) - u_{k,i} \psi_{k-1}^{(i)}\right] \\ \qquad k=1,2,\ldots,N \\ w_i \leftarrow \psi_N^{(i)} \end{cases} \qquad (16)$$

D. Diffusion LMS Solution

As mentioned in the introduction section, when more communication resources are available, it is possible to take advantage of the network connectivity and devise more sophisticated peer-to-peer cooperation rules [7]. To describe one such diffusion implementation we consider gain Fig. 1. The neighborhood of a node $\mathcal{N}_k$ is the set of nodes directly connected to it, including itself. Each node k consults peer nodes from its neighborhood and combines their past estimates $\left\{\psi_l^{(i-1)}; l \in \mathcal{N}_k(i-1)\right\}$ with its own past estimate $\psi_l^{(i-1)}$, (see Fig. 2).

The node generates an aggregate estimate $\phi_l^{(i-1)}$ and feeds it in its local adaptive filter. The strategy can be expressed as follows for LMS-type recursions [7]:

$$\phi_k^{(i-1)} = f_k\left(\psi_l^{(i-1)}; l \in \mathcal{N}_k(i-1)\right) \qquad (17)$$

$$\psi_k^{(i)} = \phi_k^{(i-1)} + \mu u_{k,i}^*\left(d_k(i) - u_{k,i}\phi_k^{(i-1)}\right) \qquad (18)$$

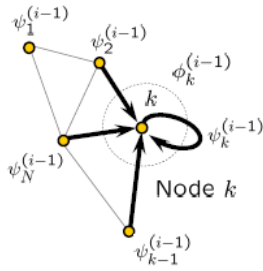for some local combiner $f_k(\cdot)$. The combiners may be non-linear or even time variant.

Fig. 2.   A network with diffusion cooperation strategy.

III. Proposed Algorithm

To deal with the mentioned conditions, it is necessary to obtain an estimate of each sensor's observation noise. To do this, we consider the equation (1) again. If we repeat the diffusion LMS algorithm (i.e. eq. (17)) Ls times (where Ls is a suitably chosen integer), it is possible to have a primary estimate of w^o. It must be noted that this estimate of w^o is used just to obtain a primary estimate of observation noise at each sensor, and it is not the final estimate of w^o. So, in the first step we must use a diffusion LMS algorithm to . In this paper we consider a linear combiner model in which at node k, the aggregated estimate $\phi_k^{(i-1)}$ is generated by linearly combining the neighbors estimates, i.e.,

$$\phi_k^{(i-1)} = \sum_{l \in \mathcal{N}_k} c(k,l)\psi_l^{(i-1)} \qquad (19)$$

$$\psi_k^{(i)} = \phi_k^{(i-1)} + \mu u_{k,i}^* \left(d_k(i) - u_{k,i}\phi_k^{(i-1)} \right) \qquad (20)$$

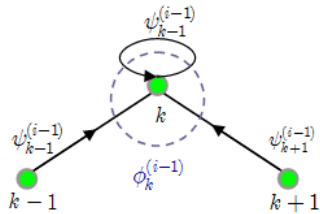Here we select a *ring* topology for the sensor network as shown in Fig. 3.

Fig. 3.   The ring topology of sensor network.

According to this topology we have

$$\mathcal{N}_k = \{k-1, k, k+1\} \qquad (21)$$

since in the first iterations $i \leq Ls$ we have not any information about observation noise we select

$$c(k,l) = \frac{1}{3} \qquad (22)$$

Now by substituting the (22) in (19) and (20), the observation noise at each sensor can be estimated as

$$n_k(i) = d_k(i) - u_{k,i}\psi_N^{(Ls)}, \quad i = 1, 2, ..., Ls \qquad (23)$$

we denote by $\psi_k^{(Ls)}$ the estimate of w^o in the ith iteration in Nth node, so we will have:

$$\psi_N^{(Ls)} = \left. \psi_k^{(i)} \right|_{k=N, i=Ls} \qquad (24)$$

In iteration $Ls + 1$, the information $n_k(i)$, $i = 1, 2, ..., Ls$, are sent to the Nth node. Having $n_k(i)$ in the N node, we calculate the variance of observation noise of the kth sensor by

$$\underset{k=1,2,...,N}{g_k} = \left(\frac{1}{Ls} \right) \sum_{i=1}^{Ls} n_k(i). \qquad (25)$$

$$\tilde{\sigma}_k = \sum_{i=1}^{Ls} \left(n_k(i) - g_k \right)^2 \qquad (26)$$

As $\tilde{\sigma}_k$ increases, the reliability of d_k decreases, so there is inverse relation between $\tilde{\sigma}_k$ and sensor's reliability. Keeping this fact in mind, we define

$$s_k = \frac{1}{\tilde{\sigma}_k}, \; k = 1, 2, ..., N \qquad (27)$$

In the next step, we define the following parameter

$$\lambda_k = \max\{s_k\}, \; k = 1, 2, ..., N, \qquad (28)$$

Finally, we define the step-size of the our diffusion LMS algorithm as

$$\mu_k = \mu_{\text{glob}}\lambda_k^{-1}s_k \qquad (29)$$

Now, we have an estimate of variance of observation noise for each sensor. So it is possible to further enhance the performance of linear combiner (19) by adjusting the $c(l,k)$ according to the estimated variances $\tilde{\sigma}_k$. To do this for sensor k, we define

$$x = \tilde{\sigma}_{k-1}, \;\; and \;\; y = \tilde{\sigma}_{k+1}, \;\; and \;\; \rho = \frac{x}{y} \qquad (30)$$

Again because of inverse relation between $\tilde{\sigma}_k$ and sensor's reliability we set

$$c(k,l) = \begin{cases} \frac{2}{3(1+\rho)} & l = k-1 \\ \frac{1}{3} & l = k \\ \frac{2\rho}{3(1+\rho)} & l = k+1 \end{cases} \qquad (31)$$

Using (29) and (31), the diffusion LMS algorithm can be written as

$$\phi_k^{(i-1)} = \sum_{l \in \mathcal{N}_k} c(k,l)\psi_l^{(i-1)} \qquad (32)$$

$$\psi_k^{(i)} = \phi_k^{(i-1)} + \mu u_{k,i}^* \left(d_k(i) - u_{k,i}\phi_k^{(i-1)} \right) \qquad (33)$$

In Fig. 4 the block diagram of the proposed algorithm is provided. it is obvious that in the proposed algorithm both step size parameters μ_k and combiner coefficients $c(l,k)$ cooperate in LMS update equation. Since these parameters are calculated according to the sensor data, so it is expected that the proposed algorithm provide a better performance. In the following section, we will present the simulation results of different algorithms.
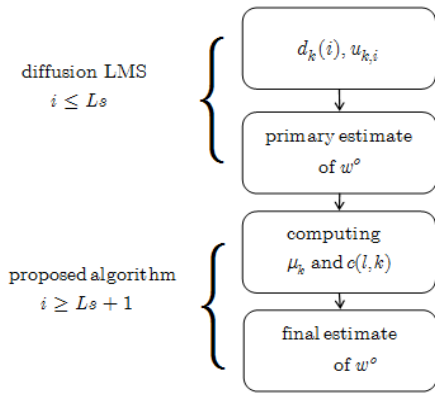
Fig. 4.   The block diagram of the proposed algorithm.

IV. SIMULATION RESULTS

In this section we present the simulation results of the proposed algorithm and compare it to other algorithms such as the IDLSM of [4] and the diffusion-LMS of [7]. To this aim we consider a network with $N = 20$ nodes and Gaussian regressors with $R_{u,k} = I$. We further assume that the variances of observation noise for sensors vary from $\sigma_v^2 = 10^{-3}$ to $\sigma_v^2 = 10^{-1}$. The curves are obtained by averaging over 100 experiments with $\mu_{\text{glob}} = 0.01$ and $M = 4$. To compare the performance of different algorithms we use the mean-square error (MSE) criteria which is defined as follows

$$\text{MSE} \triangleq E \left| e_k(\infty) \right|^2 \qquad (34)$$

The performance of the proposed algorithm is highly dependent on the value of Ls, since it determines how $\psi_k^{(Ls)}$ is close to w^o. In Fig. 5 the performance of the proposed algorithm for $Ls = 30$ in comparison with other algorithms is shown.
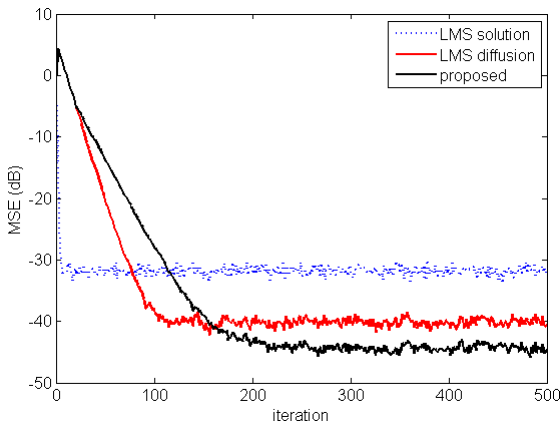
Fig. 5.   The performance of the proposed algorithm ($Ls = 30$)

As Fig. 5 shows, this value of Ls is not a suitable choose. In fact Ls must be chosen when the IDLMS algorithm (which is used to a primary estimate of w^o) is in its steady state position. According to our simulation results , $Ls = L/10$ (where L is

the total number of iterations) is a good choice. In Fig. 6, the performance of proposed algorithm for $Ls = 150$ is shown.
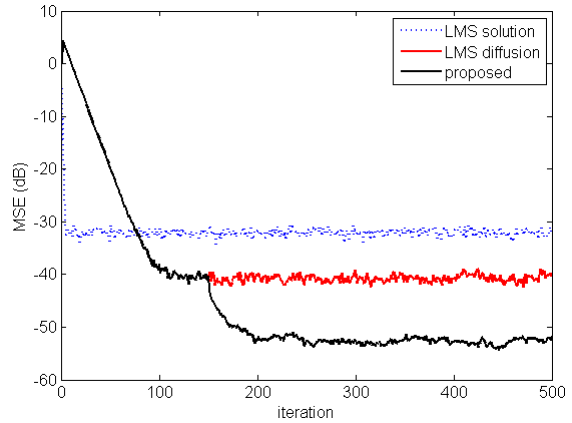
Fig. 6.   The performance of the proposed algorithm ($Ls = 150$)

As it is clear from Fig. 6, the proposed algorithm for a $L = 150$ iteration has better performance than IDLMS and diffusion LMS in a sense of steady state estimation error. The step size for different sensors (i.e. vector μ_k) is shown in Fig. 5.
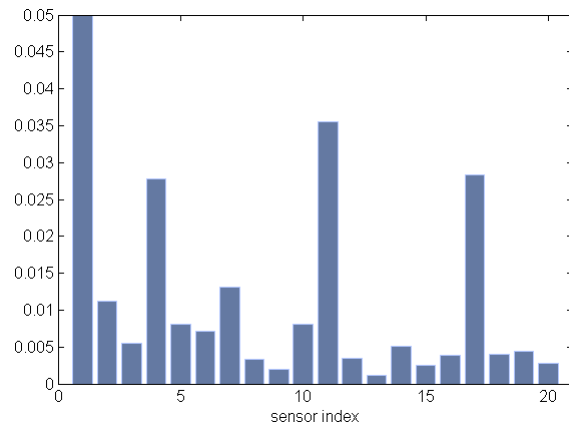

Fig. 7.   The step-size for different sensors in the proposed algorithm

Although it is possible to obtain better estimation of by choosing a $Ls > L/10$, but it must be noticed that the primary estimate of w^o is used just to obtain a primary estimate of observation noise at each sensor and it is not the final estimate of w^o. In the Fig. 8, the performance of the proposed algorithm for different Ls values is shown.

V. CONCLUSION

In this paper we considered the problem of distributed adaptive estimation over sensor networks in a more realistic scenario where different variance for observation noise is assumed for sensors in the network. Our the proposed method is divided into two phases: I) Estimating each sensor's observation noise variance and II) using the estimated variances to obtain the desired parameter. We enhanced the diffusion LMS
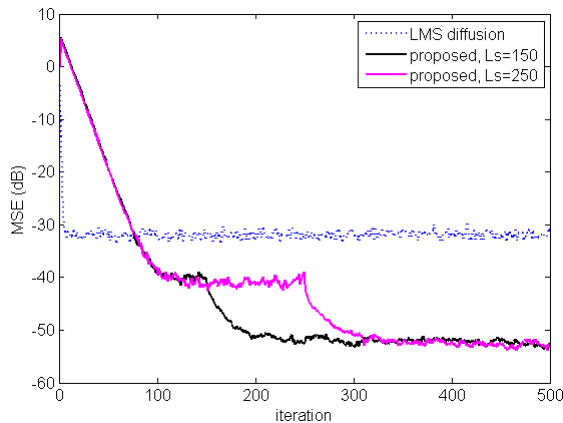
Fig. 8. The performance of the IDLMS and proposed algorithm for different Ls values.

implementation with linear combiner model. In the proposed algorithm, the step-size parameter the coefficients of linear combiner are adjusted according to estimated observation noise variances. As our simulation results show, the proposed algorithm has a better performance of diffusion LMS algorithm in the same condition.

ACKNOWLEDGMENT

REFERENCES

[1] D. Estrin, G. Pottie and M. Srivastava, Intrumenting the world with wireless sensor networks, *Proc. IEEE ICASSP*, pp. 2033-2036, May 2001.

[2] D. Bertsekas, A new class of incremental gradient methods for least squares problems, *SIAM J. Optim.*, vol.7, no. 4, pp. 913-926, Nov.1997.

[3] C. Lopes and A. H. Sayed, Distributed adaptive incremental strategies: Formulation and performance analysis, *Proc. ICASSP'06*, Toulouse, France, vol. 3, pp. 584-587, May 2006.

[4] C. Lopes and A. H. Sayed, "Distributed processing over adaptive networks, *Proc. Adaptive Sensor Array Processing Workshop*, MIT Lincoln Laboratory, MA, June 2006.

[5] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks, *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064-4077, August 2007.

[6] A. H. Sayed and C. Lopes, Distributed recursive least-squares strategies over adaptive networks, *Proc. 40th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp. 233-237, October-November, 2006.

[7] C. G. Lopes and A. H. Sayed, Diffusion least-mean-squares over adaptive networks, *Proc. ICASSP'07*, Honolulu, Hawaii, vol. 3, pp. 917-920, April 2007.

[8] F. Cattivelli, C. G. Lopes, and A. H. Sayed, A diffusion RLS scheme for distributed estimation over adaptive networks, *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Helsinki, Finland, pp. 1-5, June 2007.

[9] C. G. Lopes, and A. H. Sayed, Steady-state performance of adaptive diffusion least-mean squares, *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, pp. 136-140, Madison, WI, August 2007.

[10] C. G. Lopes and A. H. Sayed, Diffusion least-mean squares over adaptive networks: Formulation and performance analysis, *to appear in IEEE Transactions on Signal Processing*, 2008.

[11] F. Cattivelli, C. G. Lopes, and A. H. Sayed, Diffusion recursive least-squares for distributed estimation over adaptive networks, *to appear in IEEE Transactions on Signal Processing*, 2008.

Amir Rastegarnia (S'07) was born in 1981 in Iran. He received the B.S. degree and the M.S. degree in electrical engineering from the University of Tbariz, Iran, in 2004 and 2006, respectively, where he is currently pursuing the Ph.D. degree in electrical engineering. His current research interests are theory and methods for adaptive and statistical signal processing, distributed adaptive estimation, as well as signal processing for communications. He is a student memebr of *IEEE*.

Mohammad Ali Tinati (was born in 1953 in Iran. He received his B.S. degree (with high honor) in 1977, his M.S. degree in 1978 from Northeastern University, Boston, Mass, USA, and his Ph.D. degree from Adelaide University, Australia, in 1999. He had a long affiliation with the University of Tabriz, Iran. He served as an academic member of the Faculty of Electrical Engineering since 1979. His main research interests are biomedical signal processing and speech and image processing.

Azam Khalili was born in 1982 in Iran. He received the B.S. degree from K.N Toosi university of technology, Tehran Iran, and the M.S. degree from the University of Tbariz, Iran, in 2005 and 2007, respectively, where he is currently pursuing the Ph.D. degree in electrical engineering. Her current research interests are statistical signal processing, distributed adaptive estimation, as well as speech processing.