# S.E.E.D - SPEECH ENABLING EQUIPMENT FOR DUMB

*Giving words to the speechless using Artificial Intelligence*

**Farha Khan[1], Kuntal Barua[2]**

[1](M.Tech Scholar, ECE Dept., Lakshmi Narain College of Technology, Indore, M.P.)
[2](Asst. Professor, CSE Dept., Lakshmi Narain College of Technology, Indore, M.P.)

***Abstract- The most predominant part of communication is Speech. Unluckily the Dumb people are deprived of this blessing. Because of the disorders in the vocal chords or problems in that certain portion of brain which control speech, they are unable to form the sound of the words they are willing to speak. Artificial intelligence plays an important role in helping dumb people speak in the way normal people do. In this paper we explored a new device SEED which makes use of Electromagnetic Articulography to produce the voice based on the lips movement made by the dumb person. The process begins with sensor analyzing the lips movement and calculating the distance between various reference points assigned on the mouth. This distance is then mapped into corresponding vectors representing digital images, these digital images are then sampled and these digital neural pulses are then transmitted to the neuromorphic chip. The neuromorphic chip produces impulses by comparing the incoming digital neural pulses with the standard information contained by each neuron in the chip. The produced impulses are then fed to a 'Spoken phonetic transcription software'- PHONWEB, which is the genuine voice of the 'dumb person'.***

***Keywords- Articulography, neuromorphic, EMA, AG500, AG501, artificial neurons, neural network, KNN, RBF, Manhattan distance, L.SUP.***

## 1. INTRODUCTION

A wide variety of devices are already available for people suffering from physical challenges like blindness, deafness and dumbness. However the devices for dumb people limited and not used widely, the main reasons behind it are the large size and unaffordable costs of the available devices making them impractical for day to day use. Besides these the researches mainly focus on gesture recognition techniques which are underwhelming for wide commercial usage. In this paper we are exploring a new gadget SEED based on the fundamentals of artificial intelligence deploying Electro Magnetic Articulography (EMA)[1].

## 2. PRINCIPLE OF THE ELECTROMAGNETIC ARTICULOGRAPHY: EMA

EMA is a technique for recording the movement of speech articulators. The horizontal (X) and vertical (Y) positions of multiple sensors attached to the vocal organs can be recorded as a function of time. Three EM- field generating transmitter coils that are attached to a head mount produce an alternating magnetic field (M.F.). The EM-field generating coils can be considered as antennas for very low frequency radio waves (however the waves are not solely considered as radio waves, because

they differ in some characteristics from regular radio waves). The M.F. induces an alternating current in sensors that are attached to the articulators (tongue, lips, jaw, etc). The magnitude of the signal (current) is inversely proportional to the X (horizontal) and Y (vertical) distance of each sensor from the transmitter coils. A receiver coil is a kind of dipole and therefore has five degrees of freedom; these are the three X, Y and Z coordinates and the 2 angles that describe the alignment of the dipole. Thus, we can say that the signal varies not only with the distance between transmitter and receiver coil, but also with the angle between transmitter and receiver axis.

### 2.1. OPERATING PRINCIPLE OF 3- D ELECTROMAGNETIC ARTICULOGRAPHY[2]

As already discussed above, the state of the miniature receiver coil is described by five variables representing the position in the 3-D coordinate system and the rotation angles relative to it. In order to determine the position in a 3-D area, all five values must be known. Furthermore, the system should be capable of scaling all directions with same efficiency. This is realized by the spherical placement of six transmitter coils. Thus between every two transmitter coils there is a right angle. Each coil indicates a value, therefore, mathematically we can say that, we have six equations consisting five unknowns. When aligning the transmitter coils one should consider that the induced voltage becomes zero when the transmitter and receiver axis are perpendicular. In this case, no information is available concerning the distance between transmitter and receiver.

Little connector coils are placed on and inside the dumb person's mouth. Along with the supporting software, the whole process right from the beginning i.e. setup-measurement- data analysis enables the user to access useful knowledge regarding movement of articulators.

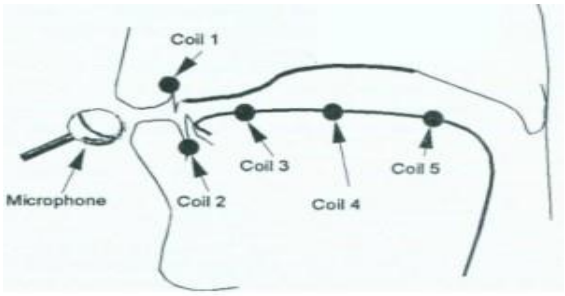


**Fig.1.** *Mid-sagittal view of experimental setup showing location of EMA coil*

#### 2.1.1. 3-D EMA DEPLOYING AG-500 SENSOR

Up to 2012, one of the most widely used component for EMA in three dimensions is the AG500 [3]. The AG500 allows the determination of the positions and orientations of up to 12 sensors (receiver coils) attached onto articulators and reference points, moving inside a registered spherical mass of radius 150 mm and immersed in a superposition of six alternating MFs, generated by one transmitter coil each. Figure2 shows the final coil mounting structure for the 3-D-EMA in a side view with 4 of the 6 transmitter coils visible. The structure consists of 2 triangles with a transmitter coil on each vertex. The pre-amplified signal is digitized by a National Instruments Multi-I/O-card, which is part of the Receiver-PC and written to disk. It is planned to use this second PC for real-time data preprocessing and to use a third PC for position-calculation and analysis of the data. So at the moment we are using up to three PC's on a Windows-NT platform to perform data acquisition and analysis and a fourth PC to move the receiver coil on a

predefined path. It is obvious that the device is not very handy at this stage of development, but there are multiprocessor PC's available by now so, various tasks involved in processing is simplified using IC CM1K. Results showed that errors in position tracking are due to numerical issues and cannot be attributed to external interference.

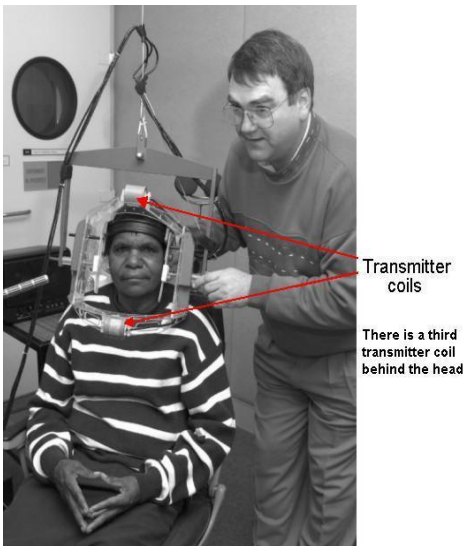


**Fig.2.** *a real image showing the positions of transmitter coils.*

### 2.1.1. RECENT DEVELOPMENTS IN THE 3D- EMA DEPLOYING AG-501 SENSOR

Several independent studies have pointed out the presence of some anomalies in the positions retrieved by the AG500. Errors up to 2 mm were observed for a pair of sensors attached to the jaw dental plate during some speech functions. Further experiments with the receiver coils rotating with constant velocities, across a circumference in a horizontal plane centered in the registered mass shows a 0.5 mm mean error, in the worst case, which can extend up to 5 mm and dispense around a median of 2.5 mm. Combating all position retrieval and other sorts of problems recently, a newer model of EMA, the AG501, has been introduced [4]. In SEED we are utilizing this EMA model. The spatial distribution of system's transmitter coils has been completely redesigned and shows much advancement. It keeps an account of the positions and the orientations of up to 24 sensors with an operating sampling rate of 250 Hz. In contrast to the previous model, instead of 6 transmitter coils we are using 9; furthermore, the coils stand above the registered mass, and the plastic cube structure.

**Fig.3.***The new Carstens System*

## 3. THE CM1K CHIP

CogniMem Technologies' CM1K is the first ASIC version of the CogniMem neural network. It features 1024 neurons functioning simultaneously, capable of learning and recognizing patterns of up to 256 bytes in a few microseconds. The two non-linear classifiers (RBF and KNN) supported by the CM1K can classify patterns while coping with inexplicit data, unknown events, and changes of surroundings and working circumstances. The parallel architecture of the CM1K allows daisy-chaining many chips together to increase the size of the neural network in increments of 1024 neurons. The less number of pins and less power dissipation makes it an excellent complement chip for smart sensors and cameras

### 3.1. ARCHITECTURE OF CM1K CHIP

CogniMem has a very simple architecture: it is a chain of identical neurons operating in parallel. A neuron is an associative memory which can autonomously compare an incoming pattern with its reference pattern. During the recognition of an input vector, all the neurons communicate briefly with each other (for 16 clock cycles) to find which one has the best match. In addition to its register-level instructions, CM1K integrates a built-in recognition engine which can receive vector data directly through a digital input bus, broadcast it to the neurons and return the best-fit category 3 microseconds later. The CogniMem neural network is controlled by a set of 6 global registers and 8 neuron registers. The neurons can learn and recognize input vectors autonomously. If several neurons recognize a pattern (i.e. —fire), the response of all of them can be retrieved in increasing order of distance (equivalent to a decreasing order of confidence). The first response is therefore the category of the first neuron with a distance register equal to the smallest distance. It is called the best match category.

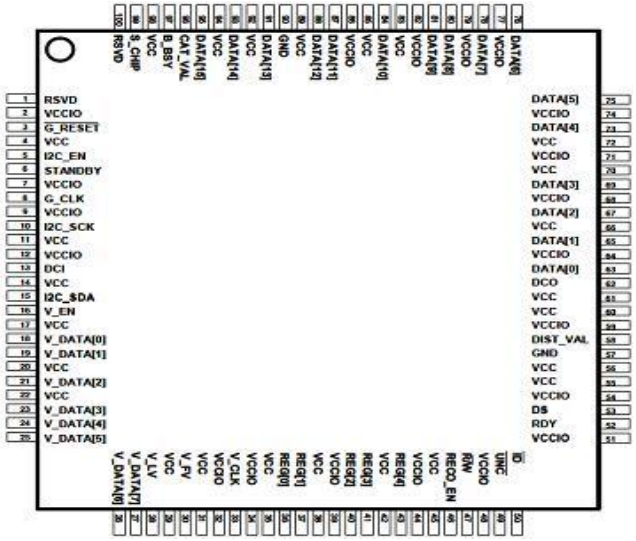


**Fig.4.** *CM1K pin diagram*

CogniMem comes with a recognition engine which is optimized to return this response when a vector is received directly on the digital input bus of the chip. During a learning session, the neurons' behavior can be tuned to be more or less conservative by changing the possible range of their influence fields. Fig 4. shows CM1K pin details.

### 3.2. TRAINING THE CM1K CHIP

The dumb's lip movements are detected and then digitized in AG500 and then it is being fed to CM1K as input. The input after decoding is compared with the neurons present in the chip via parallel 12C bus. The whole process is shown in fig.5. CM1K has the choice of two non-linear behaviors of recognition:

- K- Nearest Neighbour- KNN (pattern recognition)
- Radial Basis Function- RBF (machine learning)

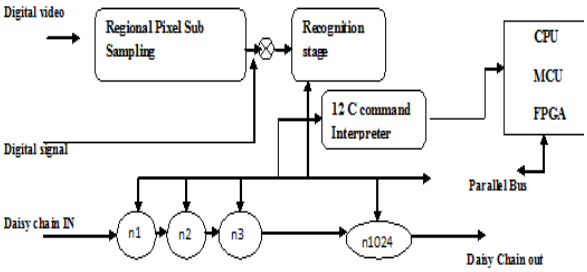


**Fig.5.** *inside the CM1K chip*

A neuron evaluates the distance between the incoming vector and the vector stored in its memory. If the distance falls within its current influence of field it returns a positive group of alphabets, here we need to store the phonetics of each alphabet in the neuron memory.

The neuron calculates two types of distances:

i. Manhattan distance (L1)(sum of all distances between incoming vector V and the pattern vector P in each neuron)

i.e..,

$$D_{L1} = \Sigma |V_i - P_i|$$

ii. L.SUP(maximum component of distances between V and P)

i.e.,

$$D_{LSup} = \text{Max} | V_i - P_i|$$

The state of neuron in the chain can be idle, RTL or committed. It is defined by the status of its Daisy Chain In (DCI) and Daisy Chain Out (DCO) lines. The DCO of neurons rises if its DCI is high and its category register shows a value other than 0. As a result, the neurons' commitment is generated automatically as examples are trained and retained. The RTL neuron continue to move along until no more idle neuron is available in the chain.

The neural network is composed of M+1+N neurons.

- M = committed neurons holding a reference pattern and a category value.
- 1 = ready to learn (RTL) neuron.
- N = idle neurons.

### 3.2.1. KNN ALGORITHM

The ***k*-nearest neighbor algorithm ( *k*-NN)** is a method for classifying objects based on closest training examples in the feature space. The parallel architecture of the NeuroMem chip makes it the fastest candidate to retrieve the K closest neighbors of a vector among ANY number by (1) calculating distances in parallel and (2) sorting them in increasing order autonomously. One of the key features of a NeuroMem network is that the programming and latency of its operations remain the same whether your network is composed of a single CM1K chip or a chain of number of CM1K chips.

#### 3.2.1.1. LOADING THE TRAINING EXAMPLES

The training examples can be any number of vectors composed of up to 256 bytes. A vector can be a series of measurements with different dimensions and characterizing a population of objects. In some cases, the 256 bytes can be samples of the temporal evolution of a signal, or the spatial distribution of pixels, and more. The training examples be loaded sequentially into the neurons using the Save and Restore (SR) mode of the CM1K chip. Under this mode, the neurons are passive and writing a neuron register takes one system clock cycle. The following diagram describes the simple sequence of commands loading the training examples to the neurons. Its execution time is proportional to the number of examples and independent of the architecture of the neural network.

Initially the Write Forget command resets the category of all the neurons to 0. It is necessary to execute this command otherwise the examples will be appended with those ones already loaded in the existing committed neurons.

The neurons are then set to the Save and Restore mode by setting bit 4 of the Neuron Status Register (NSR) to 1 and the first neuron of the chain becomes the "Ready-To-Load". The N bytes of the first example are loaded through a series of Write Component. After the $N^{th}$ component, the Write Category assigns a value to the Category register of the neuron (default is 1). The latter becomes "Committed" and the next neuron in the chain becomes "Ready-To-Load".

Once the loading of all examples is done, the network is returned to its default Learn and Recognize mode by setting bit 4 of the NSR back to 0.

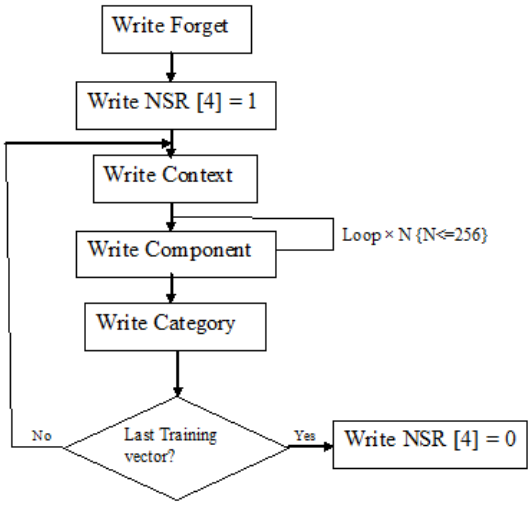


**Fig.6.** *Process of loading the training examples to neurons*

### 3.2.2. RBF ALGORITHM

Here the data is compared to the neuron memory and the output purely depends on only distance and its field influence. The shorter the distance, the greater it's field of influence and so the particular neuron having the shorter distance will be selected for producing the output.

Text synthesis:
Now the digital impulses from the sensors are being fed to CM1K. Now the vector input is compared with reference vectors in the respective neurons and the best match neuron information is the output of the dumb's lip movement which is then fed to the system to view it in the form of text. From there the required process can be performed.

## 4. VOICE CONVERSION

The output of the CM1K chip is then fed to 'text to voice conversion software' such as verbose or phonweb. With verbose one can read aloud or save spoken text to mp3 files.

## 5. IDENTIFIED PROBLEMS AND FUTURE PROSPECTS

- As we had already discussed above, there were some anomalies recognized with sensor position tracking when AG500 was the only commercially available sensor. The sensor positions obtained by AG500 sensor in some cases were afflicted by external perturbations, and some numerical issues. To combat those anomalies, in this paper we used AG501 sensor instead of AG500. However the accuracy of AG501 is by far superior, although further enhancements and trials are necessary, in order to completely rule out the numerical issues.
- In case of CM1K chip, the response of the firing neurons is ordered per increasing distance and for the same distance per increasing category. The CM1K chip does not provide for the subsequent sorting per identifier. If one or more of these neurons have the same distance and same category, the readout of the Distance register followed by the Category register will exclude them all at once from the next search and sort. This means that the neurons with the same distance and same category will be accounted as one and produce incorrect results.
- Besides above mentioned limitations, SEED is very expensive ($100,000+) and very complicated to use, mainly because of the size of the equipment. Although, with the use of CM1K chip the size is considerably reduced as compared to earlier articulograph equipments which required 3 to 4 computers for processing the incoming inputs. But still further enhancement is required to make this equipment more compact and easy to handle and carry. With every day enhancements in the technology of electronic equipments, there is a hope that in future the size of the transmitter coils as well as whole system

can be further reduced, so that the equipment can be made mobile.

- There are many sources of potential error: e.g. sensors are inaccurately placed or become unstuck or break. The size or shape of the vocal organs also influences trajectory movement.

## 6. CONCLUSION

Artificial Intelligence is playing a vital role in making better the lives of speechless people, who are deprived of the glorious blessing of voice. We look forward to make this wonderful equipment SEED more practical and affordable to common man so that in future we can completely eradicate dumbness from this planet.

## 7. REFERENCES

***[1]Jonathan Harrington, Speech physiology October 2001, Macquarie University, Sydney Australia.***

***[2]Three dimensional articulography: a measurement principle, J Acoust Soc Am. 2005 Jul; 118(1):428-43.***

***[3] Stella, M., Bernardini, P., Sigona, F., Stella, A., Grimaldi, M. & Gili Fivela, B., "Numerical instabilities and three-dimensional electromagnetic articulography", J. Acoust. Soc. Am., 132 (6), 3941-3949, 2012.***

***[4] Kroos C., "Measurement Accuracy in 3D Electromagnetic Articulography (Carstens AG500)" in Proceedings of the 8th Seminar on Speech Production, edited by R. Sock, S. Fuchs, and Y. Laprie, (INRIA, Strasbourg, France), pp 61-64, 2008.***