

Controlling a prosthetic arm with a throat microphone

Elena Mainardi, Angelo Davalli

Abstract—The aim of the present paper is to illustrate the phases of design and application of an innovative input source for an EMG upper limb prosthetic arm: a laryngophone, otherwise called throat microphone (t – mic). In the last years several different input sources were explored, from the implantable myoelectric sensors to the mechanomyographic sensors. The idea of controlling a prosthesis with vocal commands is quite recent but seems to be promising in helping users to better control their devices, improving the quality of their life.

I. INTRODUCTION

An upper limb myoelectric prosthetic arm is an aid that tries to give a chance of a better quality of life to disabled people. It tries to give back some of the natural and fundamental functions of a physiological human arm, even if the movements that is able to perform are not so deeply similar to those of a natural arm.

To control such a device, several ways are possible. The more traditional one is nowadays the EMG (electromyographic) control, which is based on EMG signals extracted from surface electrodes of user’s arm or forearm, while the most simple technique is using buttons or switches when the electromyographic activity of patient muscles is not so good or clear. Lately, in the last decade more complex ways were explored to widen the range of possible input sources for the controller of a prosthetic arm, so neuro cortical control [1], foot control with wireless wearable insoles [2], control with implantable myoelectric sensors (IMES) [3], with MMG sensors (mechanomyographic) [4] and ultrasonic sensors [5] were investigated, even if is not clear if these techniques are really used by patients in their everyday life or if they are just interesting theoretical contributions in the wide field of prosthetic arm control.

All these techniques start from the assumption that the prosthetic motion is directly “linked” to the human motion, the source being both the emg activity or the foot motion or something else. Since the most common control scheme for an upper limb prosthetic arm is a sequential control (where signals or switches are used to change control from one degree of freedom to another) , it follows that all these techniques have the same problem: when the patient has to perform a complex task, formed by a predetermined and precise sequence of movements, he has to do a precise sequence of contractions/movements, always remembering

which motor is selected in every instant of time. This is not so simple as one can believe, especially when there are more degrees of freedom (typically three for a transhumeral or a shoulder disarticulated patient: the flexion/extension of the elbow, the prono/supination of the wrist and the opening/closing of the hand). In figure 1 we can see a typical sequential control with three EMG sources.

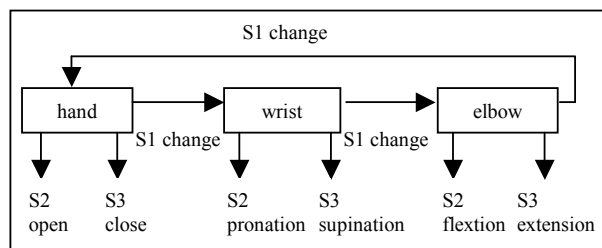


Fig. 1 Sequential control of an EMG prosthesis with three EMG sources: source 1 (S1) controls the action of changing the selected motor, sources 2 and 3 (S2, S3) control the motions of the different motors.

Let consider the following case: a patient has to bring a bottle and pour water into his glass. The sequence of contractions that he must do, in the case of sequential control of the motors and thinking about three sources of emg signal, is illustrated in table 1:

TABLE I		
Change motor →	Flexion, CW rotation, opening → pectoral muscle	Extension, CCW rotation, closing → scapular muscle
Step	Action	Muscular contraction
1	Opening the hand	Pectoral
2	Closing the hand (to bring the bottle)	Shoulder blade muscle
3	Selecting the wrist	Trapezius
4	Rotate CCW (pour water into glass)	Pectoral
5	Rotate CW (taking back the bottle in vertical position)	Shoulder blade muscle
6	Selecting the elbow	Trapezius
7	Selecting the hand	Trapezius
8	Opening the hand (release bottle)	Pectoral
9	Closing the hand (to take the glass)	Shoulder blade muscle
10	Selecting the wrist	Trapezius
11	Selecting the elbow	Trapezius
12	Flexing the elbow (to bring the glass to the mouth)	Pectoral

From the table above we can understand that even if the motion task seems to be very easy, the patient has to do a precise sequence of twelve contractions to perform it.

Manuscript received 14th April, 2007. This work was supported in part by the INAIL Prosthetic Centre (Vigorso di Budrio, BO, Italy)

E. Mainardi is with the Engineering Department of the University of Ferrara, Italy (phone +39 0532 974803, e-mail: emainardi@unife.it).

A. Davalli is with the INAIL Prosthetic Centre (Vigorso di Budrio, Bologna, Italy. email: a.davalli@inail.it)

If we think that now, in some cases, there is the idea to add, besides the elbow, the wrist and the hand motors, also a shoulder motor group with two motors (see [6]), one for the intra-extra rotation and one for the elevation-adduction, (so adding four possible movements, with other four sources of emg signals) it is clear that controlling the prosthetic device only with emg signals could become more and more difficult, also because the possible EMG sources located in the muscles near the amputation line are not utilizable, due to a bad or insufficient EMG activity.

For this reason we thought to another alternative input source, potentially efficient and easy to be used, and overall disconnected from the human body motion. In particular we focused of the voice control.

II. USING THE VOICE TO CONTROL A DEVICE

Nowadays using the voice to control an electronic device is a quite common process, and there are several electronic equipments that can be commanded by voice, such as telephones, surgery robots, wheelchairs ([7], [8]), military devices ([10]) and so on.

A voice recognition system is composed by an input device, typically a microphone, and an intelligent core that performs the recognition operations, which are, for the most part, software elaborations of the signal acquired from the input device. Explaining the recognition techniques is not the aim of the paper, but information can be found in [11]. In this work we used an off the shelf voice recognition system, the Voice Extreme Module produced by Sensory ([12]). It is an embedded module with several interface signals (digital I/O, UART interface, DAC, I²C lines...) which is programmed in C language by means of a development board and then can be connected to some dedicated embedded hardware. In this case, the embedded hardware is the control board of the prosthetic device, which acquire the inputs (typically three emg signals), elaborates them and performs the motor actions requested by the patient.

The voice recognition process is articulated in two different phases: the first one, called training phase, where the module is taught with the words that it must recognize, and a second phase, the standard operation, where one pronounces a word and the module compares it with the stored words and decide which of them is the most similar one. The module is programmed in voice dependent mode, so it can recognize a word only if it is pronounced by the same person that has done the training phase.

In this context, the voice command is not intended to completely substitute the traditional EMG control, but just to join it, to expand the possibilities of controlling the device, and to simplify the control process in case of complex and repetitive motion tasks.

III. USING A MICROPHONE OR USING A LARYNGOPHONE

The idea of using a voice recognition system to control a prosthetic arm is very recent.

In the literature there are just only two examples (see [13]-[15]).

Two main problems in this form of recognition are represented by the input sound level and the additive noise. Regarding the first one, level changes result from speakers speaking with different volume or from changes of orientation or distance from the microphone. Concerning the second one, we have to consider that when someone is speaking there are almost always other sources of sound (other voices, radio, TV, traffic noise etc.). If training and testing phases are carried out at the same noise level, the recogniser can cope quite well even with low signal-to-noise ratio, but if there is a mismatch this can cause severe performance degradations. Also, in a noisy environment, if the surrounding noise is not so high and the microphone is directional and very close to the mouth, a voice recognition system can work fine anyway. But when the surrounding noise become higher, with simply a traditional microphone the recognition process can become difficult. It's true that there exist some techniques to improve the quality of the recognition even in a noisy environment ([16] - [20]), but is still true that over a certain noise sound level the standard microphone will fail in its working. In these conditions, and also to keep the level of recognition process relatively simple, one can use a laryngophone, or throat microphone (t-mic).

The main difference is that a traditional "air" microphone transforms the air pressure caused by voice emission from the mouth in an electric signal, using a condenser or a coil ([21]), while a t-mic often uses a piezoelectric sensor to convert the throat vibrations into an electric signal, since the vocalization is a phenomenon produced by muscular activities of trunk, larynx, mouth and tongue. So an air microphone is affected by the environmental noise potentially present, while a throat microphone is not (or at least not in the most common cases).

For this reason a throat microphone is more robust and can work fine even if the patient is by the way with the traffic noise, or there's a phone ringing, or other people are talking, or the radio is playing music. A standard microphone, in these situations, would often fail.

This is the reason why, after having designed a traditional voice recognition system (VRS) to control a prosthetic arm [15], we decided to design also a VRS with a laryngophone (or t-mic, i.e. throat microphone, or also p-mic, i.e. physiological microphone).

T-mics are under investigation in the field of voice recognition in noisy environment ([17], [21]), while they have been used from several years in military applications and in the radio-communication field.

The structure of the system is illustrated in figure 2.

IV. INTERFACING THE THROAT MICROPHONE

Standard preamplified condenser microphones are well known devices, and several systems that use voice utilities

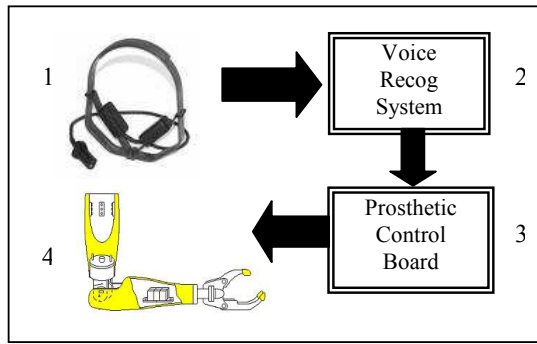


Fig. 2 The structure of the system

are designed to be interfaced with such devices. The same is for the Voice Extreme module.

Different is for a throat microphone, that is not preamplified and is not a standard device. T-mic are often used in radio communication, but every radio device has different characteristics and so every t-mic is different from the others. Our t-mic is a T05 model by Midland.

First of all, the piezoelectric sensor has to be powered. Then, since the way the sensor receive power is the same that it uses to convey the voice signal, the line has to be DC-decoupled with a condenser. After that the signal has to be amplified and rearranged in impedance. At the end, it can be connected to the input line of the voice recognition module. We use a trimmer to adjust the gain of the amplifier, so the sensor can be calibrated on the vocal vibration intensity of the patient.

V. THE CHOICE OF THE WORDS

If we think about a voice control that simply reproduces the EMG sequential control (a vocal command for every action that can be controlled by a muscular contraction) we would need to store four words, one to change the selected motor, two to control the motions of the motors (like sources 2 and 3 in figure 1) and the last one to stop the action.

The words used with the voice recognition system and the standard microphone were the Italian “avanti” and “retro” (“forward” and “back”) to select the action, “cambio” (“change”) to change the selected motor and “stop” to stop the action.

Of course, changing input device and being the t-mic different from an air microphone, we had to test other words to select the best set. In every test a comparison between the response of a microphone and the throat microphone was made. Of course we had to think about words related to the actions of the device to be controlled (something like “go”, or “action”) and quite different from each others, so the voice module doesn’t get confused ([22]).

In figure 3 there are some results.

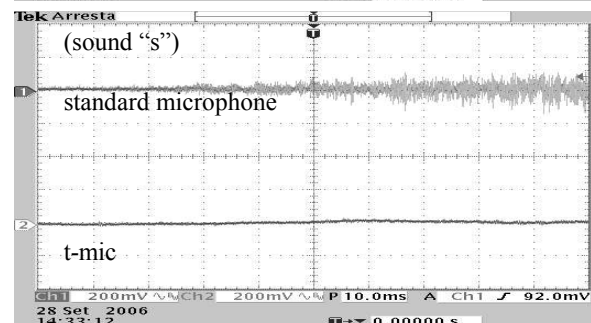
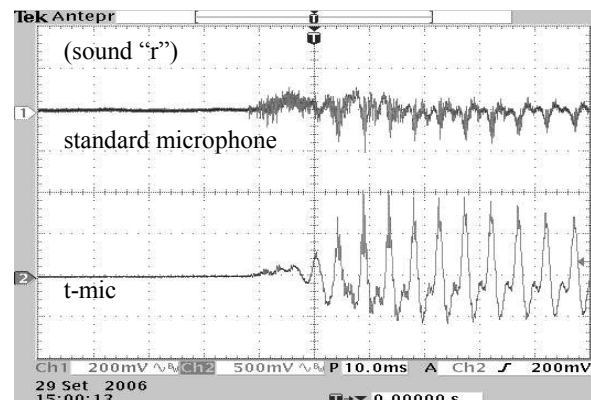
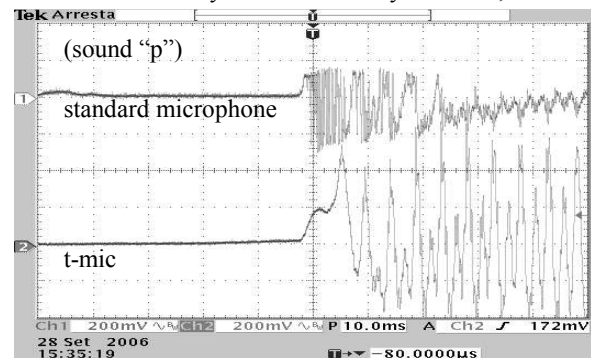
It can be seen that the signal coming out from the throat microphone is very different from the one produced by the standard microphone. In particular the letter “s” is almost zero for a t-mic (trace 2 in the figure where there are two traces) because it causes almost no vibration of the larynx,

instead is revealed by the standard microphone because the sound associated with this letter produces a little air compression. It can also be noted that the letter “p” is revealed before with the standard mic, and after a few milliseconds with the t-mic, because the t-mic is more sensible to the vocalic sound associated to the end of the consonant, respect to the sound of the consonant itself (when we pronounce “p”, in point of fact we pronounce “p” plus a vocalic sound, and is this latter one that the t-mic reveals better). Also we can notice that the sound “s” and “su” are very different, so as “u” and “su”. This is due to the fact that every sound is affected by the previous pronounced sound. At the end, we can note that the letter “u” is very well revealed by the t-mic.

After several tests, we selected the Italian words “su” and “giu” (“up” and “down”) to control the actions related with the movements of the prosthesis, “cambia” (“change”) to control the action of changing the selected motor and “stop” to stop the motion.

These four words were selected for the following reasons:

- the sound “u” is very well revealed by the t-mic, and the



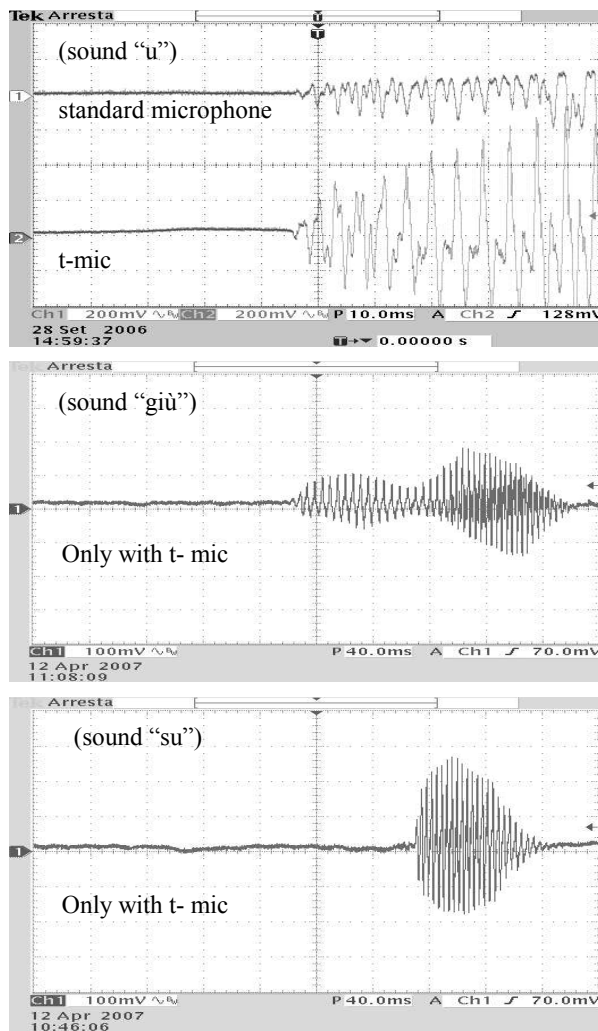


Fig. 3 The signals of the standard microphone and the t-mic

sounds "su" and "giu" are very different from each others, so the voice recognition system is able to separate them;

- "cambia" is formed by letters well revealed by the t-mic, and also the sound of "cambia" is longer and very different from "su" and "giu", so the voice recognition system doesn't get confused
- the sound of "stop" is impulsive and different from both "su" and "giu".

After the basic idea that the same actions that a patient can do with the EMG control can be done with voice control (e.g. moving a limb or changing the selected motor), there is a more challenging process that can be useful for people with prosthesis: using a single word to perform a complex task, executing a sequence of several movements.

In this case we thought about two possible tasks (just to test the system, but of course the only limit to the number of tasks that can be done in this way is the memory size of our control board): the first one is taking a bottle and pouring its content in a glass, while the second one is taking a glass and drink (bringing the glass to the mouth).

The only difference between the voice control of the first case and the one regarding the execution of sequence of motions is that in this last one we chosen to put the voice module in a state of "continuous listening" until the user pronounce a key word that awakes it. After having pronounced this word, the user has to say one of the words that are associated with a sequence of actions.

The words were selected with the same criteria of the first case, and are "protesi" ("prosthesis") to awake the module, "versa" ("pour") to command the first task, "bicchiere" ("glass") to command the second task.

VI. THE CONTROL WITH THE T-MIC

As is said before, the voice control is not substituting the standard EMG control of the prosthetic device, but widens the range of possible inputs. A patient can decide to switch from the EMG signal to the voice control simply doing a particular and known sequence of contractions, and can do the opposite passage pronouncing a particular word associated to the switching off of the voice module and the reactivation of the EMG signals.

When the voice control is active, the motion associated with the command words are predetermined and stored in the control board, so the prosthesis executes always the same motions and the patient has to interact with his aid. For instance, for a right prosthesis, when the selected task is taking a bottle and pouring water in a glass, the sequence starts opening the hands, so the patient has to move the prosthesis near the bottle, then the prosthesis closes the hand (e.g. bringing the bottle), then the wrist turns left (pouring water), then it turns right (so the bottle comes back to the vertical position), then the hand opens (releasing the bottle).

This could seem very strange, but in fact of reality is very useful for real patients. Besides this, the interaction between the patient and the predetermined actions is necessary because the control of the device is always in an open loop scheme. The prosthesis is not able to adapt its behaviour to the situation or the context it works in because it doesn't have sensors like cameras or others. But this interaction is the same that patients do every time they use their prosthesis, even without the voice control, because they always adjust the position of their remnant limbs to perform many tasks (for instance when they write they take the pen with the prosthetic hand and then go back and forth with the shoulder to move the hand). Just because this interaction is very natural, it has been used in this voice control, even because otherwise it would have been impossible to realize it.

Of course one important thing to be considered is setting the right timings for all the actions executed by the prosthesis. This is a phase that must be performed with the same care of the phase of choosing the best places to locate the EMG sensors on the muscles of the patient. But, anyway, once that the timings and the motions have been stored in the control board, if we have chosen that the wrist has to turn 90° to pour water in the glass and the bottle is more or less full of water, the patient would rotate his trunk or his shoulder to adjust this angle and so on. These are operations

that patients do every time, and are more complex to be explained that to be really performed by patients.

VII. RESULTS

The system have actually been tested on able-body people. The percentage of words well recognized is near 97%.

We have tested the system with the two tasks mentioned before.

Table 2 shows the comparison between the time that is necessary to complete the tasks for able-body people, able-body people using a standard upper limb EMG prosthesis with EMG control and able-body people using a standard upper limb prosthesis with t-mic control.

	Time for TASK 1	Time for TASK 2
Able-body people	5	5.5
Able-body people with EMG	40	40
Able-body people with t-mic	15	15

Finally we also tried to discriminate between a standard and a throat microphone in term of robustness in a noisy environment. We used the pc to generate a white noise with different sound levels and a fonometer (NA29E by Rion) to measure the sound level, testing the voice recognition module with the two different input devices. We noted that with a 75 dB of white noise, the voice recognition system with a standard microphone positioned 5 cm from the mouth, aligned with the mouth, doesn't work anymore, while the same system with the t-mic still works fine.

VIII. CONCLUSION AND FUTURE DIRECTIONS

The promising results are viewed as just a first step in the application of a throat microphone to control a prosthetic device. Second phase of the project will be testing the system on real patients with prosthesis, improving the performances of the circuit, maybe integrating it in the same board of the prosthetic electronic controller, and also thinking about substituting the t-mic with a wireless t-mic, so that the patient does not have to wear uncomfortable wires that run between the board and the t-mic.

ACKNOWLEDGMENT

The authors thank engineers Davide Masi and Carlo Colturato.

REFERENCES

- [1] Schwartz, Andrew B.; Taylor, Dawn M.; Tillery, Stephen I. Helms, "Extraction algorithms for cortical control of arm prosthetics", *Current Opinion in Neurobiology* Volume: 11, Issue: 6, December 1, 2001, pp. 701-708
- [2] M. C. Carrozza, A. Persichetti, C. Laschi, F. Vecchi, R. Lazzarini, V. Tamburrelli, P. Vacalebri, P. Dario, "A Novel Wearable Interface for Robotic Hand Prostheses", *Proceedings of the 2005 IEEE 9th International Conference on Rehabilitation Robotics* June 28 - July 1, 2005, Chicago, IL, USA
- [3] R. F. Weir, P. R. Troyk, G. DeMichele, D. Kerns, "Technical Details of the Implantable Myoelectric Sensor (IMES) System for Multifunction Prosthesis Control", *Proceedings of the 2005 IEEE EMBS Engineering in Medicine and Biology 27th Annual Conference* Shanghai, China, September 1-4, 2005
- [4] J. Silva, T. Chau, A. Goldenberg, "MMG-Based Multisensor Data Fusion for Prosthesis Control", *Proceedings of the 25th Annual International Conference of the IEEE CMBS Cancun, Mexico* September 17-21, 2003
- [5] T. Tanaka, S. Hori, R. Yamaguchi, M. Q. Feng, S. Moromugi, "Ultrasonic Sensor Disk for Detecting Muscular Force" *Proceedings of the 2003 IEEE International Workshop on Robot and Human Interactive Communication* Millbrae, California, USA, Oct. 31 -Nov. 2, 2003
- [6] Troncossi M., Parenti-Castelli V., Davalli A., "Mechanical design of a prosthetic shoulder mechanism for upper limb amputees," *ICORR 2005, IEEE 9th International Conference on Rehabilitation Robotics: Frontiers of the Human-Machine Interface*, June 28 - July 1, 2005, Chicago, Illinois
- [7] Obando, Marcelo A.; Payne, John H., The future application of the robotic arm (Automatic Endoscopic System for Optimal Positioning or AESOP) with voice recognition in sinus endoscopic surgery, *Operative Techniques in Otolaryngology - Head and Neck Surgery* Volume: 14, Issue: 1, March, 2003, pp. 55-57
- [8] Mazo, M.; Rodríguez, F.J.; Lázaro, J.L.; Ureña, J.; García, J.C.; et al., "Electronic control of a wheelchair guided by voice commands", *Control Engineering Practice* Volume: 3, Issue: 5, March, 1995, pp. 665-674
- [9] B. Beek, E. P. Neuberg, D. C. Hodge, An Assessment Of The Technology Of Automatic Speech Recognition For Military Applications, *Ieee Transactions On Acoustics, Speech, And Signal Processing*, Vol. Assp-25, No. 4, August 1977
- [10] J.P. Woodard, E. J. Cupples, Selected 'Military Applications of Automatic Speech Recognition Technology, *Communications Magazine*, IEEE Volume 21, Issue 9, Dec 1983 Page(s):35 - 41
- [11] Keller, E. (Ed.). (1994) *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art and Future Challenges*. Chichester, UK: John Wiley (380 pp.).
- [12] Voice Extreme Module, © 2001 SENSORY, INC. ALL RIGHT RESERVED, available at www.sensoryinc.com
- [13] Chih-Lung Lin; Shao-Cheng Wang; Hen-Chong Wu; Shuenn-Tsong Young; Maw-Huei Lee; Te-Son Kuo; A speech controlled artificial limb based on DSP chip, *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE EMBS*
- [14] K Towers, K. Barnes, C. Wallace, "Voice recognition for prosthetic control case study", *Proceeding of MEC 2005*, August 15 - 19, 2005, Fredericton, New Brunswick, Canada
- [15] E. Mainardi, S. Banzi, A. Davalli, (In Italian) "Applicazioni Di Riconoscimento Vocale per Protesi Mioelettriche di Arto Superiore", A.N.I.P.L.A. Biosys, June 2005, Milan (Italy)
- [16] Furui, Sadaoki, Toward Robust Speech Recognition and Understanding, *The Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology* Volume: 41, Issue: 3, November, 2005, pp. 245-254
- [17] M. Graciarena, H. Franco, K. Sonmez, H. Bratt, "Combining Standard and Throat Microphones for Robust Speech Recognition", *IEEE Signal processing letters*, vol. 10, no. 3, March 2003
- [18] H.W. Ruehl, S. Dobler, J. Weith, P. Meyer, A. Noll, H.H. Hamer and H. Piotrowski, "Speech recognition in the noisy car environment", *Speech Commun*, 10(1), 11-22 (Feb. 1991)
- [19] B. H. Juang, "Speech recognition in adverse environments", *Comput. Speech and lang*, 5, 275-294 (1991)
- [20] R. Haeb-Umbach, D. Geller and H. Ney, „Improvements in speech recognition for voice dialing in the car environment", *proceedings of ESCA-ETRW Workshop on Speech Processing in Adverse Condition*, Cannes-Mandelieu (France), Nov 1992, pp.203-206
- [21] L.E. Kinsler, A.R. Frey, A.B. Coppens, J.V. Sanders, "Fundamentals of Acoustic, third edition", Jhon Wiley&Sons Edition, 1980
- [21] A. Shahina, B. Yegnanarayana, "Language identification in noisy environment using throat microphone signals", *Proceedings of ICISIP 2005*
- [22] Heungkyu Lee, Hanseok Ko, "Competing models-based text-prompted speaker independent verification algorithm", *Speech Communication* Volume: 48, Issue: 1, January, 2006, pp. 28-44