ACOUSTIC SENSOR FOR VOICE WITH EMBEDDED PHYSIOLOGY

Michael V. Scanlon

Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783-1197
mscanlon@arl.mil

Military Sensing Symposium
November16-19, 1999
North Charleston, SC

ABSTRACT

The Army Research Laboratory (ARL) is developing sensor technology to monitor the soldier’s voice and physiology by using enhanced acoustic sensors. The physiological sensor consists of liquid or gel contained within a small, conformable, rubber bladder or pad that also includes a hydrophone. This enables the collection of high signal-to-noise ratio cardiac, respiratory, voice, and other physiological data. The pad also minimizes interference from ambient noise because it couples poorly with airborne noise. It is low cost and comfortable to wear for extended periods. When the sensor pad is in contact with a patient's thorax, neck, or temple region, sounds can be immediately and continuously monitored. This can aid in the remote assessment, diagnosis, and treatment of cardiac and respiratory functions, as well as provide human stress and performance indicators such as heart and breath rates, voice stress, and gross motor indicators.

The neck sensor picks up the wearer's voice well, with fidelity sufficient to be used as a hands-free voice activation mechanism using speech recognition software, or for voice communications. The pulse is also detectable from the carotid artery, and excellent breath sounds are present as well. This attachment area is often unobstructed by other equipment or clothing, and is easy to attach quickly to a subject.

Data presented will compare a microphone held in front of a speaker’s mouth to that of a fluid sensor held in contact with the neck. Comparing the amplitudes of the voice to the non-vocalized ambient noise surrounding the voice gives approximately 40 dB SNR for the airborne microphone, and approximately 75 dB SNR for the fluid-coupled sensor. The fluid coupling represents an improvement of better than 30 dB in SNR with minimal waveform degradation, as observed by the similar spectrograms and by listening to the data through headphones.

The ability of body-coupled sensors to detect voice in high background noise was investigated. A gel sensor was attached to a speaker’s neck. Positioned in front of the person’s mouth was a boom microphone. Data presented shows simultaneously collected breath and voice data before, during, and after a speaking subject is submerged in a 105 dB C noise field while speaking a counting sequence repeatedly. The boom microphone did not detect any voice during the high amplitude noise. However, the counting is clearly visible in the spectrograms throughout the noise with the body-coupled gel sensor.

Form SF298 Citation Data

Report Date *("DD MON YYYY")* 00001999	**Report Type** N/A	**Dates Covered (from... to)** *("DD MON YYYY")*

Title and Subtitle Acoustic Sensor for Voice with Embedded Physiology	**Contract or Grant Number**
	Program Element Number
Authors	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Army Research Laboratory 2800 Powder Mill Road Adelphi, MD 20783-1197	**Performing Organization Number(s)**
Sponsoring/Monitoring Agency Name(s) and Address(es)	**Monitoring Agency Acronym**
	Monitoring Agency Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes	
Abstract	
Subject Terms	
Document Classification unclassified	**Classification of SF298** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** unlimited
Number of Pages 8	

Playing the data collected through headsets, a listener can clearly hear and understand the spoken words from the gel sensor, but not the boom microphone.

The results demonstrate that the gel-coupled sensor can be useful for monitoring communications as well as physiology. Unlike most medical sensor technologies that look at only one physiological variable, a *single acoustic sensor* can collect information related to the function of the heart, lungs, or it can detect changes in voice or sleep patterns, motor activity, and mobility. It also provides situational awareness clues as to how the soldier is interacting with the battlefield in relation to the mission.

1.0 BACKGROUND

ARL has developed a new method to measure human physiology and monitor health and performance parameters. This consists of an acoustic sensor positioned inside a fluid-filled bladder in contact with the human body. Packaging the sensor in this manner minimizes outside environmental interferences, and signals within the body are transmitted to the sensor bladder with minimal losses. This fluid-coupling technology comfortably conforms to the human body, and enhances the signal-to-noise-ratio (SNR) of human physiology to that of ambient noise. An acoustic sensor of this type has a tremendous potential for determining soldier stress levels during performance-type tasks. An acoustic sensor system can detect changes in a person’s physiological status resulting from exertion or injuries such as trauma, penetrating wound, hypothermia, dehydration, heat stress, and many other conditions (or illnesses). Indications of a dangerous condition can be used to recommend corrective procedures or alerts to medical personnel or supervisors can be initiated. Managers can use preparedness and physiological data as a decisional aid for human resource allocations. Training leaders and participants can monitor performance levels for the presence of dangerous physiological conditions.

A sensor contacting the torso, head, or throat region picks up the wearer's voice very well through the flesh, with fidelity sufficient to be used as an auxiliary microphone for communications or hands-free voice activation mechanism. For example, the sensor would detect voice commands to allow a natural language interface to interpret the commands and initiate an action. Speech recognition software, in conjunction with this enhanced body-coupling sensor, could improve mission performance by reducing false voice commands through improved SNR in noisy environments. Because of the sensor’s ability to attenuate ambient noise, another nearby soldier's voice would not be detected as well due to losses in both amplitude and frequency content. If minimal airborne coupling did occur, the extraneous voice waveform would attenuated and distorted to the point that it would be rejected as a proper voice command. Data presented later in this report show that the physiological sensor can reject more than 30 dB of ambient noise.

Civilian technology transfer applications include clinical surveillance in convalescent and Veterans Administration homes, medical transports, hospitals, and telemedicine applications [Scanlon, patents]. Fire, rescue, and police personnel may benefit from hands free voice communications with embedded health and performance monitoring. Drivers of vehicles and aircraft could also be monitored for the onset of sleep, seizure, or heart attack.

2.0 ADVANTAGE OF ACOUSTICS

Cardiovascular activity is a function of many factors: cognitive activity, respiration, physical exercise, temperature, and chemical, emotional, and physiological factors. Monitoring heart rate only makes it difficult to interpret the resulting physiology accurately because all contributing factors may not be accounted for. A single acoustic sensor can monitor diverse physiological indicators. Quantifying more physiological indicators can result in better understanding of the entire picture. Acoustic sensing provides a low-cost, lightweight, and adaptable means of monitoring many aspects of a human (or animal) subject's physiology, as well as its interaction with the environment that may be influencing performance. Software algorithms that evaluate data from acoustic sensors can be continuously modified to monitor new parameters or the correlation between different body functions, and lends itself to data fusion with other sensing technologies.

3.0 SENSOR PLACEMENT

There are significant trade-offs to be considered for placement of sensors at different body locations. One of these is user acceptance. If the user (test subject) does not like the attachment location, sensor placement, or attachment method because it is uncomfortable, or interferes with his normal activity or abilities, it will adversely affect the test/mission and will not be useful. Also of importance is the availability or presence of an acoustic signature at a location that relates to a physiological parameter being studied. Obviously, the further the sensor is placed from the throat region, the less intelligible voice will be detected. This also is an SNR issue, because other physiology, motion, or external noise may mask the signal of interest. An example of this might be the loss of relatively quiet breath sound data detectable by a chest sensor during intense physical lifting or motion that uses chest and arm muscles.

Several different sensor configurations developed for evaluation include torso-mounted, neck and wrist attachments, and PASGT helmet headband mount. The headband attachment detects the temple pulse, breath sounds through the sinus cavities and tissues, as well as speech through bone and tissue conduction. This sensor could also be attached to a helmet headband, hat, or gas mask. The neck attachment collects excellent speech and breath sounds. The pulse is also detectable from the carotid artery. This attachment area is often unobstructed by other equipment or clothing, and is easy to attach quickly to a subject. The availability of other physiological events, such as cough, gag, wheeze, vomit, eat, drink, and swallow can be important indicators under situations of nuclear, biological, chemical, and environmental effects.

An adaptable chest harness was designed to allow flexible placement of the sensor at various points on the front of the torso, back, and sides. Changes may be necessary during tests to determine where certain physiological sounds are loudest, or where to place a sensor so as not to interfere with other equipment or hardware. A simplified chest attachment hangs from the neck by a simple band. The placement of the sensor falls just above the sternum, near the aortic valve to allow collection of good heart sounds as well as breath sounds. Voice detection at this location is only somewhat intelligible because higher-frequency components resulting from the vocal cords and mouth/sinus influence are not picked up as well. A disadvantage of a chest strap is long-term comfort because of the sensor's interaction with other equipment such as clothing, ballistic vest, load bearing equipment, and backpack.

4.0 SENSING ELEMENT

The physiological sensor system consists of a housing, gel-coupling sack with sensor embedded within, neck strap, preamplifier, and battery pack with hardwired signal egress and push to talk button. Neck band sensors are shown in figures 1 and 2. The headband sensor shown in figure 3 does not use a liquid coupling, but rather an acoustically conductive silicone rubber.

Figure 1: Gel sensor pad with shroud and strap.

Figure 2: Physiological neck assembly for voice.

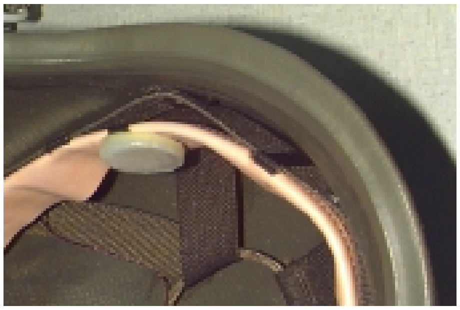

Figure 3: Rubber-coupled sensor in helmet headband.

The fluid or rubber provides acoustic impedance matching, much like high-performance clinical and industrial ultrasonic transducers that require a matching layer with controlled acoustic properties.

5.0 SENSOR TESTS

Data were collected at the side of the neck using a sensor of similar geometry to the sensor in figure 1, but water was used instead of gel as the coupling medium [Scanlon, 1998]. The test included a spoken word count from 1 to 10, then mouth breathing for the remainder of the data set. Naturally, the heartbeat is always present. The time and frequency representations are shown in figure 4.

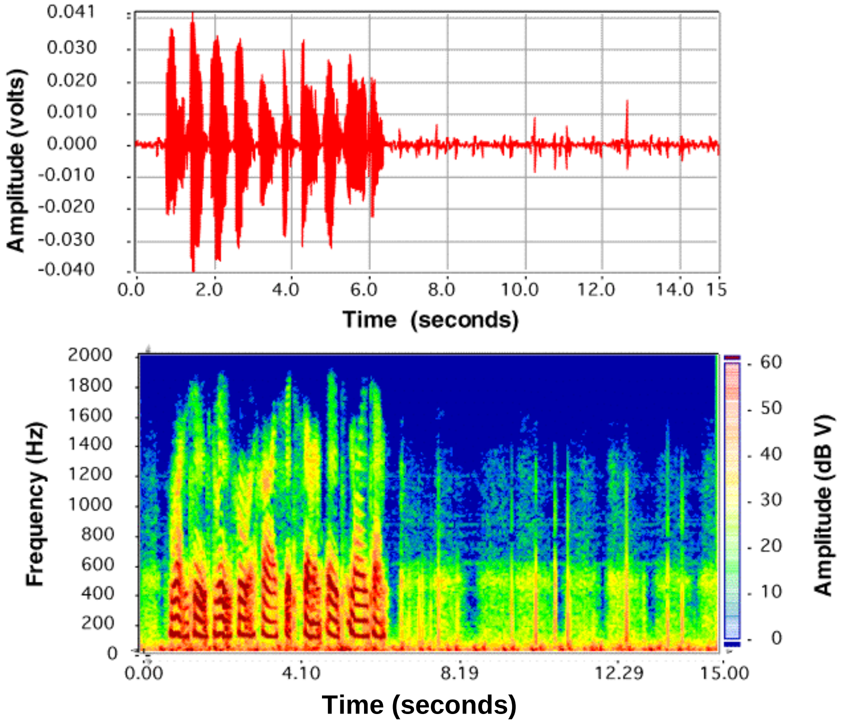

Figure 4: Fluid sensor held at throat for 1 to 10 voice count and mouth breaths.

Note in both the time-waveform and the spectrogram of figure 4 the high SNR of voice compared to the "physiological ambient noise" that includes heartbeats and breaths. The voice is so loud at the throat that the preamplifier gain must be adjusted to one of the lower gain settings to avoid amplifier saturation during speech. This excellent coupling for voice, when combined with the sensor's inherent noise immunity, could make this sensor location ideal for monitoring voice for voice-stress analysis and communications, in addition to physiology. Note also how clear the breath indications are.

6.0 HIGH SPEECH SNR

Figure 5 compares data from a B&K microphone in front of the speaker's mouth to that of a fluid-coupled physiological sensor held in contact with the neck by a strap. Data from both locations were taken simultaneously in a typical office environment. Comparing the amplitudes (dBs) of the voice to the non-vocal ambient noise surrounding the voice gives approximately 40 dB SNR for the B&K airborne microphone, and approximately 75 dB SNR for the fluid-coupled sensor. The fluid coupling represents an improvement of better than 30 dB in speech SNR with minimal waveform degradation, as observed by the similar spectrograms and by listening to the data through headphones.

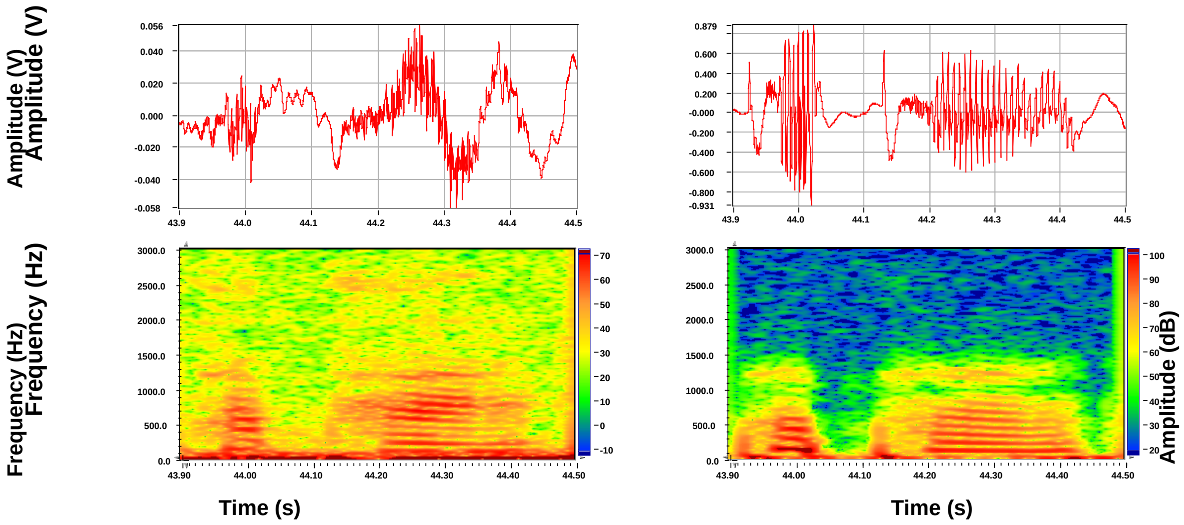

Figure 5: Comparison of spoken word “papa” taken with ambient microphone (left) and throat pad (right).

7.0 HIGH NOISE ENVIRONMENT

The detection of physiology and voice in high noise environments is very important for medical evaluation during evacuation, vehicle/aircraft operator monitoring, or voice commands in a high noise environment, such as a tactical operation center with multiple speakers. The ability of body-coupled sensors to detect physiology and reduce background noise was investigated, with results shown below. A 1-inch piezoceramic disk embedded within aqueous-couplant gel was attached to one side of a speaker’s neck. Positioned in front of the person’s mouth was a Knowles 1994 microphone, in the boom microphone configuration used previously.

Figures 6 and 7 show simultaneously collected breath and voice data before, during, and after a speaking subject is submerged in a C-weighted noise field of 105 dB (referenced to 20 micropascals, measured near the throat) noise field inside an acoustic anechoic chamber. Hearing protection was required in such a loud environment. The person wearing the sensors repeatedly vocalized a 1 to 10 count between the times of 14- and 19-s, 25- to 33-s, 65- to 71-s, and 71- to 77-s, and vocalized “105 dB” between 47- and 50-s.

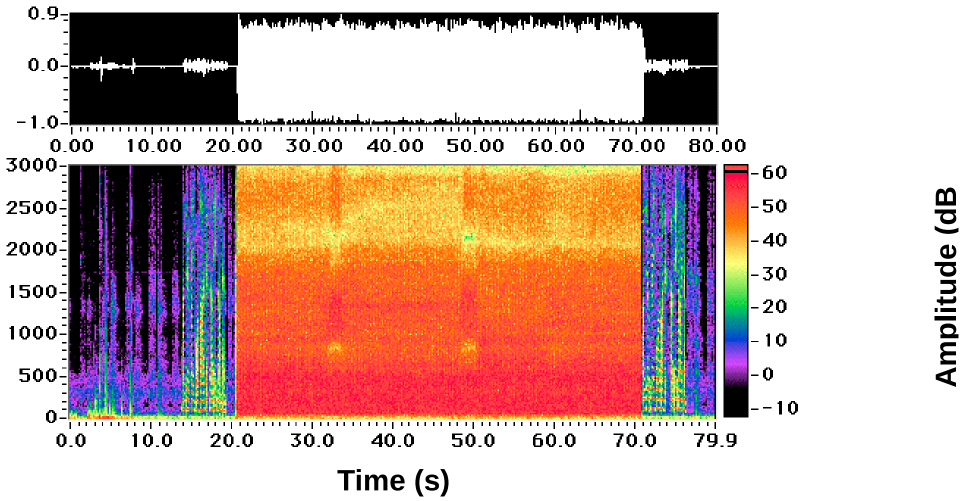

Figure 6: Boom microphone detecting voice in high noise environment (105 dB, C-weighted).

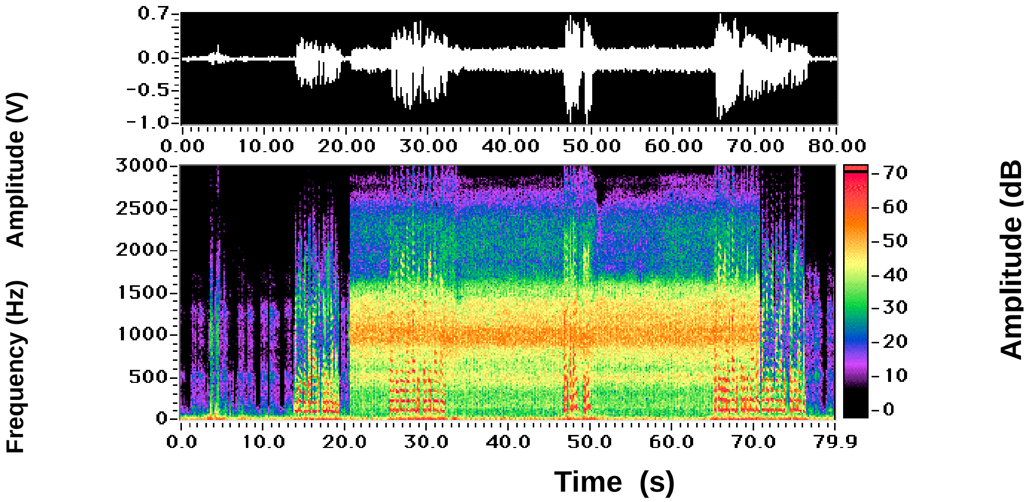

Figure 7: Gel sensor on neck detecting voice in high noise environment (105 dB C).

The boom microphone in figure 6 does not detect any voice during the high amplitude noise between 20 and 71 s. However, in figure 7, the counting is clearly visible throughout the loud noise with the body-coupled gel sensor. Playing the data collected through headsets, the listener could clearly hear and understand the spoken words from the gel sensor in 105 dB noise, but could not discern the presence of any speech in the boom microphone data.

8.0 WORD RECOGNITION IN NOISE ENVIRONMENTS

To further evaluate the capabilities of this sensing technology, speech recognition at various noise levels was studied using the sensor shown in figure 2. The United States Military Academy's Foreign Language department faculty and staff used cadet volunteers segregated into training and testing groups to evaluate 50 multi-word phrases in speech SNRs of 10-, 3-, and 0-dB SPL [Bass, Scanlon, Mills, and Morgan]. The spectrally-stable noise broadcast through speakers in front of the subject was a looped sound-bite taken from a recording made inside an M1 tank turret traveling at 25 kph. A boom microphone in front of the mouth and a physiological microphone contacting the neck were simultaneously recorded by a PC sound card during the spoken phrases. Training data from each sensor and each noise level were used to develop Hidden-Markov Model (HMM) phoneme-level speech models for each of the two sensors. When these HMMs were applied to the remaining test data, the physiological microphone clearly demonstrated superior detection of words in high noise environments.

Word accuracy results were calculated using word loop language models developed specifically for each sensor. Results indicate that at 10 dB speech SNR the airborne boom mic scored 51 percent compared to the physiological mic's 67 percent, at 3 dB SNR the airborne scored 13 percent to the physiological 50 percent, and at 0 dB SNR the airborne scored 0 percent to the physiological 40 percent. These data clearly indicate that the standard airborne microphone and its associated word recognition model fails completely in high noise environments, while the physiological microphone with its word recognition model performed very well in the same environment. Sentence loop models were used to calculate the occurrence of correct phrase selection. Results show that at 10 dB SNR both mics scored approximately 99 percent, but at 3 dB SNR the physiological scored 99 percent compared to the airborne 61 percent, and at 0 dB SNR the physiological scored 96 percent compared to the airborne 41 percent.

9.0 CONCLUSIONS

Acoustic sensors provide a low-cost, lightweight, noninvasive, and adaptable means to monitor many aspects of a soldier's health and activity. Unlike most medical sensor technologies that look at only one physiological variable, a *single acoustic sensor* can collect information related to the function of the heart, lungs, and digestive tract or it can detect changes in voice or sleep patterns, other activities, and mobility. The airborne mic's 41 percent correct phrase recognition is abysmal compared to the physiological sensor's 96 percent. Reliably detecting phrases like cease fire, fire at will, enemy target at 235 degrees, etc. is absolutely necessary for the safety of our soldiers and the success of the mission at hand. The physiological sensor has demonstrated exceptional capabilities for the detection of voice in high noise environments. The enhanced coupling and effective ambient noise rejection create very high SNR for speech and physiology. This is important in almost every military and civilian application. Acoustics can provide invaluable clues to help understand the interrelations between the soldier's physiology, the task at hand, and the surrounding environment.

10.0 LITERATURE CITED

Bass, J., Scanlon, M., Mills, T., and Morgan, J., *Getting Two Birds with One Phone: An acoustic sensor for both speech recognition and medical monitoring*, November 2, 1999, Acoustical Society of America, Columbus, Ohio.

Scanlon, M. V., *Acoustic Sensor for Health Status Monitoring*, Proceedings of the 1998 Meeting of the IRIS Specialty Group on Acoustic and Seismic Sensing,.Volume II, pp. 205-222.

Scanlon, M. V., *Sudden infant death syndrome (SIDS) monitor and stimulator*, May 1996, U.S. Patent 5,515,865; *Motion and sound monitor and stimulator*, Nov. 4, 1997, U.S. Patent 5,684,460; *Acoustic monitoring sensor*, December 29, 1998, U.S. Patent 5,853,005; International patents pending.