ADP014016

TITLE: A Surface Vibration Electromagnetic Speech Sensor

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Multi-modal Speech Recognition Workshop 2002

To order the complete compilation report, use: ADA415344

The following component part numbers comprise the compilation report:
ADP014015 thru ADP014027

A Surface Vibration Electromagnetic Speech Sensor

Jonathan L. Geisheimer, Eugene F. Greneker, Scott A. Billington, Ittichote Chuckpaiwong

Abstract—As researchers continue to improve speech in noisy environments, more interest is being placed on sensors with modalities that can be fused with traditional acoustic sensors. The standard literature has shown that electromagnetic sensors can be used to detect glottal motion. Also, accelerometers placed on the throat and nasal areas have been used to detect skin surface vibrations corresponding to speech and that data has been used for noise reduction. The Georgia Tech Research Institute (GTRI) is transitioning a 24 GHz radar technology originally used for non-contact vital signs monitoring to a technology able to measure surface motion on the order of microns, which can detect skin surface vibrations corresponding to speech. The radar has been shown to measure the same motion as accelerometers using electromagnetic waves. This paper describes the theory and preliminary work in developing a surface vibration electromagnetic speech sensor to be used for noise reduction in conjunction with acoustic sensors.

***Index Terms*—radar, speech, noisy environments, sensor fusion.**

I. INTRODUCTION

Every time a person speaks, the acoustical pressure waves from speech couple through many parts of the body, which causes structures such as the head, neck, chest, and face to vibrate. If a hand is placed on the chest or throat when speaking, these vibrations can be readily felt. The acoustic pressure waves due to speech have been translated to mechanical vibrations. This has been confirmed by various researchers who have looked at the head and chest vibrations in signers.[1] Other researchers have detected mechanical vibrations off of the neck using contact accelerometers and have been successful in using the resultant vibration signal to cancel noise when fused with acoustic data.[2,3]

An electromagnetic-based sensor called the Glottal Electromagnetic Micropower Sensor (GEMS), developed at Lawerence Livermore National Laboratories,[4] has been used to detect internal body vibrations. This sensor uses a low power, wideband pulsed radar that is able to penetrate through the body and detect glottal movement.[5] It operates at microwave frequencies less than 3.0 GHz. In general, lower microwave frequencies will achieve better penetration into the body.

The surface vibration electromagnetic speech sensor concept uses electromagnetic waves in the millimeter wave region to measure the slight vibrations of the body on the skin corresponding to human speech, down to micron levels of motion. At the proposed operational frequency of 35.0 GHz, the electromagnetic waves pass through clothes but do not penetrate into the body as does the GEMS sensor. The radar is detecting only surface vibrations and therefore directly measures the surface skin vibration and not the internal body structures. Since the device is directly picking up speech vibrations, it will be referred to as a "radar microphone". A diagram of the concept is shown in Figure 1.

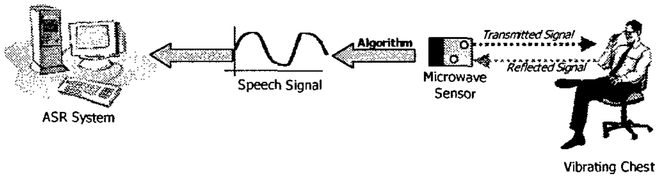

Figure 1. Radar Microphone concept

Referring to Figure 1, the radar microphone transmits a continuous wave (CW) electromagnetic signal towards the person's chest or neck area. Next, the signal is reflected back to the sensor where it is demodulated and converted to a baseband signal, sampled by an analog-to-digital converter, and then run through digital signal processing algorithms to convert the radar signal into displacement that correlates with the surface vibrations due to speech. The resultant speech signal can then be fused with other more traditional speech sensors and then passed on to an automatic speech recognition system if desired.

II. TECHNOLOGY BACKGROUND

The Georgia Tech Research Institute (GTRI) has been sensing small-scale biological motion using radar for almost 20 years, beginning with the Radar Vital Signs Monitor (RVSM). RVSM technology is able to detect both respiration and heartbeat signatures from individuals without contact. The first GTRI RVSM system was developed in the mid-1980s under sponsorship of the United States Department of Defense (DOD); a patent on the system was issued in 1992.[6] This frequency modulated (FM) radar was used as a battlefield vital signs monitor. The system was tested on soldiers wearing a chemical or biological warfare suit to allow vital signs to be monitored without opening the suit and risking contamination of the subject.[7]

J. L. Geisheimer is with the Sensors & Electromagnetic Applications Laboratory at the Georgia Tech Research Institute, Atlanta, GA 30332 USA (telephone: 770-528-7690, e-mail: jon.geish@gtri.gatech.edu).

E. F. Greneker is with the Sensors & Electromagnetic Applications Laboratory at the Georgia Tech Research Institute, Atlanta, GA 30332 USA (telephone: 770-528-7744, e-mail: gene.greneker@gtri.gatech.edu).

S. A. Billington is with the Manufacturing Research Center at Georgia Tech, Atlanta, GA 2002 USA

I. Chuckpaiwong S. A. Billington is with the Manufacturing Research Center at Georgia Tech, Atlanta, GA 2002 USA

A later version of the RVSM was developed for use in the 1996 Olympics held in Atlanta, Georgia and was addressed in a paper presented by one of the authors.[8] This system was built to monitor the heartbeat of competitors in the archery and rifle events and was able to penetrate through the heavy leather flak jackets typically used by competitors. Finally, a variant called the RADAR Flashlight was developed for use by law enforcement personnel to detect the radar respiration signature of individuals concealed behind a wall or within an enclosed space under the sponsorship of the National Institute of Justice (NIJ).[9] A picture of the latest Radar Flashlight prototype is shown in Figure 2.

Figure 2. Radar Flashlight prototype

Recent advances in the technology have increased the resolution of the sensor so it is able to detect motion on the order of microns. The associated hardware and signal processing advancements have now enabled the sensor to detected vibrational skin motion associated with speech directly off of the body.

III. SURFACE VIBRATION SPEECH SENSOR THEORY

The radar microphone is based on a phase detection technique to achieve a sensitivity high enough to pick up surface vibrations due to human speech. The key to the technique is that it does NOT use the Doppler effect or time of flight measurements common in most traditional radar designs. The key to the GTRI technique is that the sub-wavelength phase is measured with high accuracy. Motion less than the transmitted wavelength is being measured.

The radar microphone detects motion similar to a laser vibrometer, however, millimeter microwaves are used instead of light and a homodyne detection technique is being used instead of an interferometer. Typically, when electromagnetic waves are used in the context of radar or other remote sensing applications, the object of interest is moving through multiple wavelengths. If that object is moving relative to the transmitter, the received frequency will be different then the transmit frequency. This is the well-known Doppler effect. However, when an object moves less than a wavelength, such as the case in detecting chest vibrations, a different phenomenology, phase modulation, is at work.

To prove the basic fundamentals of the concept, the vibration of the chest was first recorded with a contact accelerometer and the corresponding acoustic speech was recorded with a microphone. The accelerometer was a high frequency PCB 352C68 placed on the chest and the microphone was a standard acoustical transducer. The simultaneously recorded output from the two sensors for the segment of speech "hickory dickory dock" is shown in Figure 3. The accelerometer data clearly shows many of the same characteristics as the audio signal. The radar microphone will measure the same vibrations as the accelerometer in a non-contact manner. Past research by the authors has shown that signal detected by the radar correlates well with accelerometer outputs.[10]

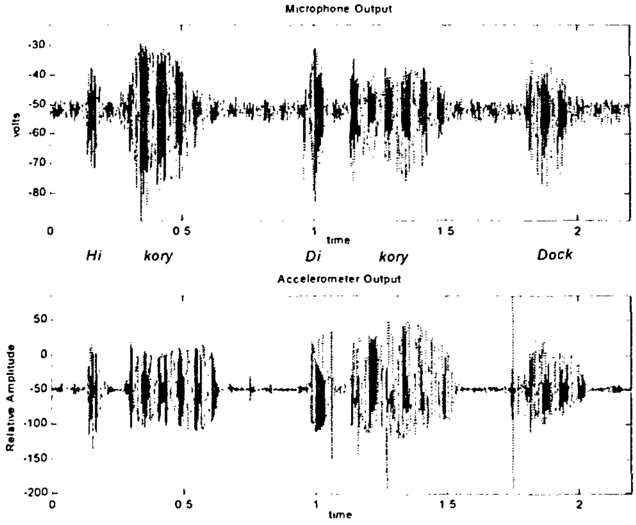

Figure 3. Simultaneous microphone and accelerometer speech data for "hickory dickory dock"

IV. PROTOTYPES

A prototype has been constructed to demonstrate the technology for a different application; however, the results are useful to show the current state of the technology as well as the promise of the radar microphone. The resulting hardware was tested using a linear motor with an optical encoder.

Figure 4 depicts the hardware configuration of the test setup. A target was attached tightly to a moving portion of a linear motor. The target surface was covered with a flat metal sheet that is used as a reflector. The radar sensor and the linear-motor encoder were set to take simultaneous measurements. The displacement from the radar sensor and the encoder were compared, consequently the radar sensor could be calibrated and compared.

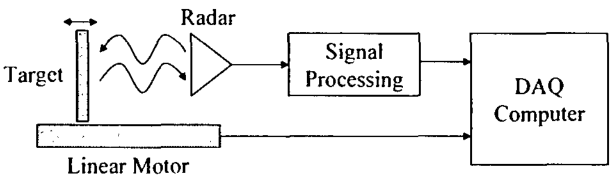

Figure 4. Radar microphone test setup

The results are illustrated in Figure 5. The top graph is a plot of both the radar sensed motion, and the ground truth motion as recorded by the encoder. It can be seen that the radar sensor was able to track actual displacement of an arbitrary motion. The residual (difference between the radar and encoder calculated displacement) on the lower graph is the difference between displacements measured by the radar sensor and the reference, or error, of the radar sensor. According to this graph, the accuracy of the radar sensor can be given to within ±1 mm over a displacement range of 50mm. Looking at smaller portions of the displacement, it can be seen that the error if often less than 0.1 mm.

Also, the residual being measured in this case is absolute displacement. Relative displacement errors have been measured down to 20 microns. Note that the residual is not randomly distributed, but a periodic function of displacement. The periodic error is caused by multipath reflections between the metal target and the metal radar hardware. Sensing of speech motion will yield significantly less multipath and distortion due to the less coherent reflecting surface.

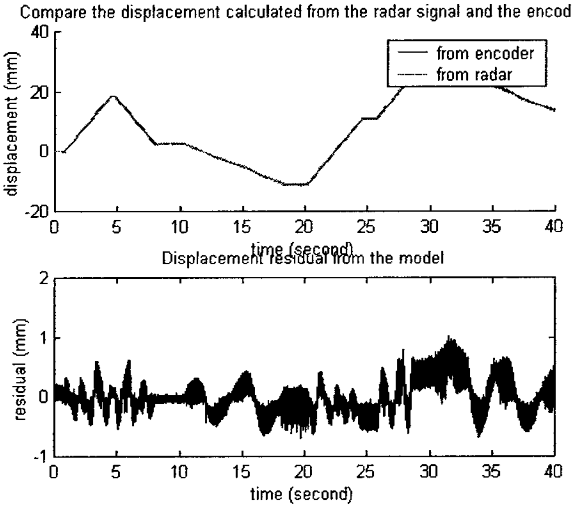

Figure 5. Example data taken from test setup

Some initial recordings have been taken using this prototype along with simultaneous acoustic recordings. After processing the radar signal, the presence of speech information is readily apparent at frequencies bellow 500 Hz and the signal correlates well with the acoustic data, however, the radar-derived speech is not yet intelligible. Increases in performance will occur both through signal processing, as well as better antenna design, which will increase the frequency response, as discussed below.

V. MODAL ANALYSIS

Critical to the successful operation of a radar microphone is the "spot size" of microwave energy illuminated by the antenna. This is critical because the sensor is measuring vibrations that are propagating along the surface of the chest. Waves with peaks and nulls are moving through the chest at different frequencies. One analogy would be the waves that move outward in water when a stone is dropped into a pond. There are peaks and nulls in the water corresponding to the propagating surface waves.

The work of Dr. Kevin Riggs at Stetson University has produced holographic images of vibratory modes in different materials. Figure 6 shows an example vibratory mode for a six inch square steel plate. The peaks and nulls on the plate are readily apparent. It is critical for accurate measurement of the vibration signal that the illumination area not detect both peaks and nulls at the same time, which may smear the output signal in the frequency domain.

Because the radar is receiving the sum of reflections from all illuminated points, the peaks and nulls could cancel each other out and distort the signal of interest. Therefore, the bandwidth of the radar microphone is limited by the antenna spot size on the chest. The smaller the spot size, the higher the frequencies that can be adequately picked up by the sensor.

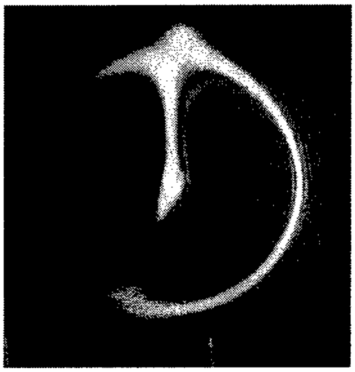

Figure 6. Example image of vibratory modes on a steel plate (K. Riggs, Stetson University)

As the standoff distance from the radar to the target of interest increases, the area illuminated by the radar beam increases, affecting the frequency sensitivity of the sensor. The spot size in centimeters vs. distance in meters for various antenna beam sizez (in degrees) is shown in Figure 7.

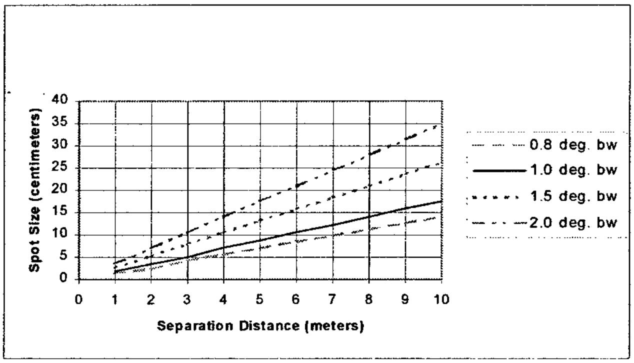

Figure 7. Spot size for given antenna beamwidths and distances

For the sensor to be viable, an antenna must be designed that projects a small spot size onto the neck, face, or chest of the person. If the application is in traditional military communications, the soldier or pilot will typically be wearing a headset, to which a sensor can be placed close to the face or neck. For larger standoffs, more exotic antennas will need to be designed. Moving the radar to a higher transmitted frequency will also enable smaller spot sizes, enhanced resolution, and improved frequency response. As advances in commercial radar technology drive prices down for operating at higher frequencies (such as 77 GHz for automobile collision control), the ability of the technology to detect high resolution speech will be improved.

VI. CONCLUSION & FUTURE DIRECTIONS

The concept of using a radar device as a surface vibration electromagnetic speech sensor has been introduced. The radar acts as a sensitive motion detector able to detect the surface vibration of skin due to speech. Testing of a 35.0 GHz sensor has shown the ability to measure motion down to microns. The next step is to take the 35.0 GHz radar sensor and record a corpus of simultaneous radar and audio data to process and compare. Signal processing algorithms will be necessary to extract speech information out of the radar data. Initial recordings using the sensor have shown the presence of speech information at 500 Hz and below in the radar signal.

REFERENCES

[1] Sundlberg, Johan, "Phonatory head and chest vibrations in singers," 127th ASA Meeting, June 6-10, 1994.

[2] Viswanathan, V., Henry, C., "Noise-immune multisensor speech input: formal subjective testing in operational conditions," ICASSP-89.,Glasgow, UK, March 1989.

[3] Viswanathan, V., Henry, C., "Noise-immune multisensor speech transduction using multiple sensors," ICASSP-85, Tampa, FL, March 1985.

[4] Burnett, G. C., et al., "The use of glottal electromagnetic micropower sensors (GEMS) in determining a voices excitation function," Proceedings of the 138th Meeting of the Acoustical Society of America, Columbus, Ohio., November 2, 1999.

[5] Holzrichter, J.F., Ng, L. C., "Speech articulator and user gesture measurements using micropower, interferometer EM-Sensors," IEEE Instrumentation and Measurement Technology Conference, Budapest, Hungary, May 21-23, 2001.

[6] J. Seals, S. R. Crowgey, and S. M. Sharpe, U.S. Patent Number 4958638, issued September 25, 1990.

[7] J. Seals, S. R. Crowgey, and S. M. Sharpe, "An electromagnetic non-contact vital signs monitor," *SOUTHCON '87 Conference Record*, 1987.

[8] E. F. Greneker, "Radar sensing of heartbeat and respiration at a distance with security applications," *Proceedings of SPIE, Radar Sensor Technology II*, Vol. 3066, pp. 22-27, Orlando, Florida, April 1997.

[9] E. F. Greneker, "Radar Sensing of Heartbeat and Respiration at a Distance with Security Applications," *Proceedings of SPIE, Radar Sensor Technology II*, Vol. 3066, Orlando, Florida, pp. 22-27, April 1997.

[10] J. L. Geisheimer, E. F. Greneker, "Neural network applications of the radarcardiogram (RCG)," *SPIE AeroSense '99*, Orlando, Florida, April 1999.