Preprint
UCRL-JC-140496-REV-1

Speech Articulator and User Gesture Measurements Using Micropower, Interferometric EM-Sensors

J. F. Holzrichter and L. C. Ng

This article was submitted to
Institute of Electrical and Electronics Engineers Instrumentation and Measurement Technology Conference, Budapest, Hungary, May 21-23, 2001

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

February 6, 2001

IEEE Instrumentation and Measurement
Technology Conference
Budapest, Hungary, May 21-23, 2001.

Speech Articulator and User Gesture Measurements Using Micropower, Interferometric EM-Sensors

J. F. Holzrichter and L. C. Ng
Lawrence Livermore National Laboratory and University of California, Davis
Livermore, California 94551 USA
Phone: +1-925-423-7454, Fax: +1-925-423-8746, Email: holzrichter1@llnl.gov

Abstract ***–Very low power, GHz frequency, "radar-like" sensors can measure a variety of motions produced by a human user of machine interface devices. These data can be obtained "at a distance" and can measure "hidden" structures. Measurements range from acoustic induced, 10-micron amplitude vibrations of vocal tract tissues, to few centimeter human speech articulator motions, to meter-class motions of the head, hands, or entire body. These EM sensors measure "fringe motions" as reflected EM waves are mixed with a local (homodyne) reference wave. These data, when processed using models of the system being measured, provide real time states of interface positions or other targets vs. time. An example is speech articulator positions vs. time in the user's body. This information appears to be useful for a surprisingly wide range of applications ranging from speech coding synthesis and recognition, speaker or object identification, noise cancellation, hand or head motions for cursor direction, and other applications***

Keywords ***– Radar, Microwaves, Speech, Coding, Microphones***

I. INTRODUCTION

Recent studies using micro-power radar-like sensors (see Fig. 1) have shown that human (i.e., animate) speech articulator motions (see Holzrichter et al. [1]) and inanimate mechanical vibrations can be measured in real time as their corresponding acoustic sounds, such as speech or musical instrument sounds, are produced. Initial work also showed that simple, non-spatially located measurements of speech articulators can provide information on a wide variety of motions associated with speech sound production—such as tissue movements in the glottal region, jaw, tongue, soft palate, lips and other areas. Similarly, characteristics of vibrating mechanical structures such as musical instrument strings, and vibrating plates are easily measured. Larger motions of a use's body parts are also being measured using "multiple fringe" counting, together with fractional fringe techniques appropriate for homodyne radar-like sensors.

For this paper, three primary application modes are consi d-ered: 1) small scale movements where Δx is $<< \lambda/4$) intermediate scale where $\Delta x \sim \lambda/4$; and 3) where $\Delta x > \lambda/4$. Because the electronics in these sensors do not work as well at low frequency (e.g., near DC) as they do at higher frequencies (e.g., > 5 Hz, and commonly > 60 Hz), and because of the "cluttered" environment in which these applications commonly take place (i.e., inside the human head or on the desk top), applications have concentrated on those structures that reflect EM waves at rates significantly greater than 5 Hz. Experiments to date have concentrated on measuring specially "targeted" structures such as the human glottal system, the jaw/tongue system and specially modulated EM wave reflectors. Modulated reflectors, employing electronic or mechanical means, are used for calibration or for attachment to a body location to enable specifically targeted motion measurements.

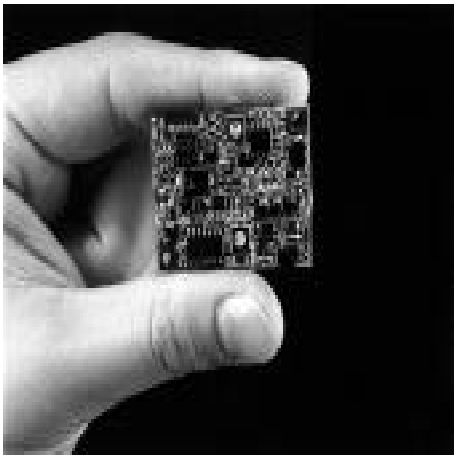

Figure 1. This microwave sensor transmits nominal 10 cycle pulse trains at frequencies ranging from 900 MHz to 5 GHz. It uses a homodyne receiver mode, radiates < 0.3 mwatts of power into 4π sr, and uses internal filtering to detect motions cycling at > 5 Hz. Two patch antennas, 1.5 cm x 1 cm, are used, one to transmit and the second to receive the EM waves. With a high-gain antenna, similar devices operating at 4 GHz have a > 50M range.

II. HOMODYNE SENSORS

The homodyne field disturbance EM sensor works as an EM wave interferometer by comparing the reflection of a transmitted wave against a local (phase reference) wave. As the targeted reflecting surface of the system moves, the phase of the reflected wave varies with respect to the stationary local wave. A signal associated with the product of the two waves is detected by a mixer, is integrated, amplified, and filtered (see Fig. 2). The detection system measures position changes (as small as 2 micrometers) versus time, within specific frequency bands ranging from <5 Hz to 7 kHz. A homodyne response of signal versus position, for an EM sensor designed to sense glottal tissue motions at 70-7 kHz, is shown in

Fig. 3. The sensor response (see Fig. 3) was measured using a 100 cm^2 Cu plate target, vibrating at 100 Hz with an amplitude of about 200 micrometers, that was moved to a new location for each data point, see Burnett [3]. This particular curve approximates the derivative of the "standard" cosine homodyne function [4], but shows a fall off with target-sensor range, and shows some positive/negative signal value asymmetries, probably due to mixer-diode asymmetries. When these data are used to determine a distance in a non-unity dielectric system, (e.g., in human tissue with $\varepsilon \approx 50$) the effective wave path in air, which is about 7 times that of the equivalent tissue path, must be taken into account. The thin arrow in Fig. 2, at 70 mm, shows the effective position on the sensitivity curve of the rear windpipe wall of a subject. The large, numbered arrows illustrate the three different modes of using these sensors, as described in the introduction: 1) small distance, 2) medium distance, and 3) large distance. It is common, but not used in the work described here, to obtain a second output signal from a homodyne sensor, that is "in quadrature" to the first "normal" signal output. Using this second signal, a second set of homodyne fringes, like those shown in Fig. 3 but shifted by 1/4 wave, can be obtained; and using these, the direction of motion can be determined [4].

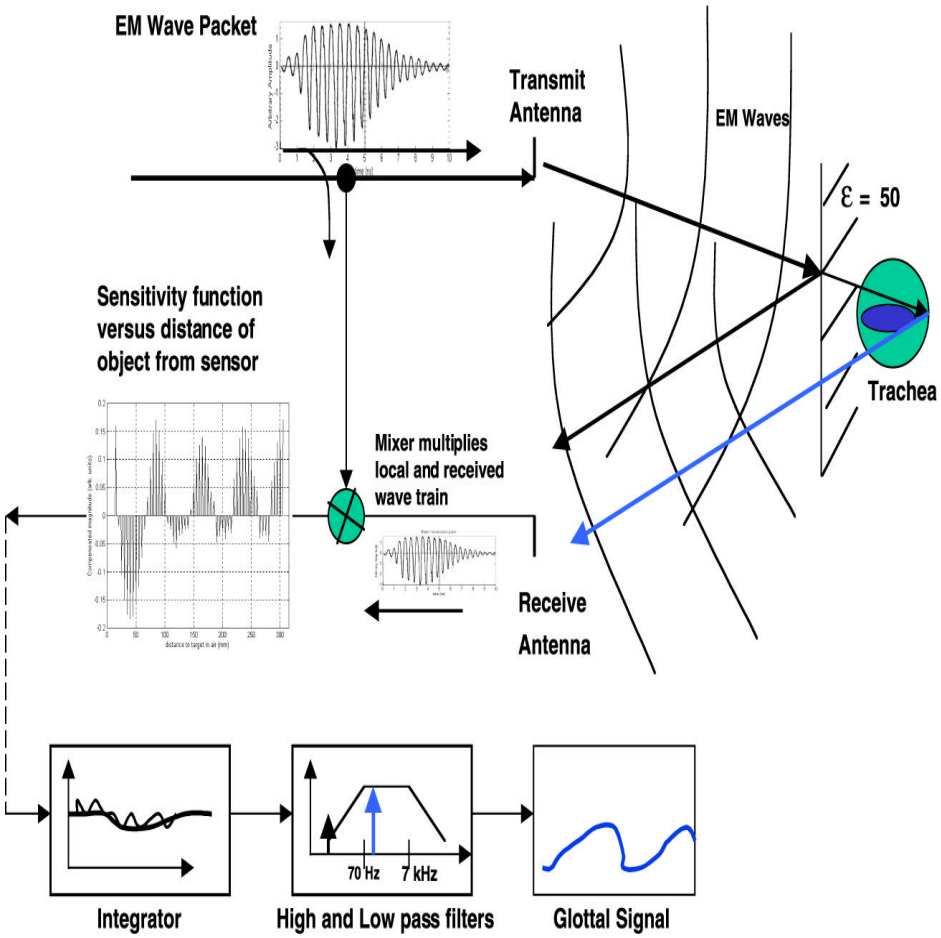

Figure 2. Typical homodyne circuit when used in a small-signal, field-disturbance mode to measure glottal tissue motions. A wave packet is generated by a gated oscillator (upper left) and is both transmitted to a target (e.g., human glottal and tracheal regions) and to a mixer. Upon reflection, the EM signal is directed to a diode mixer, then the "zero"-frequency envelope (from the mixer) is integrated, filtered, and processed to described interface distances versus time.

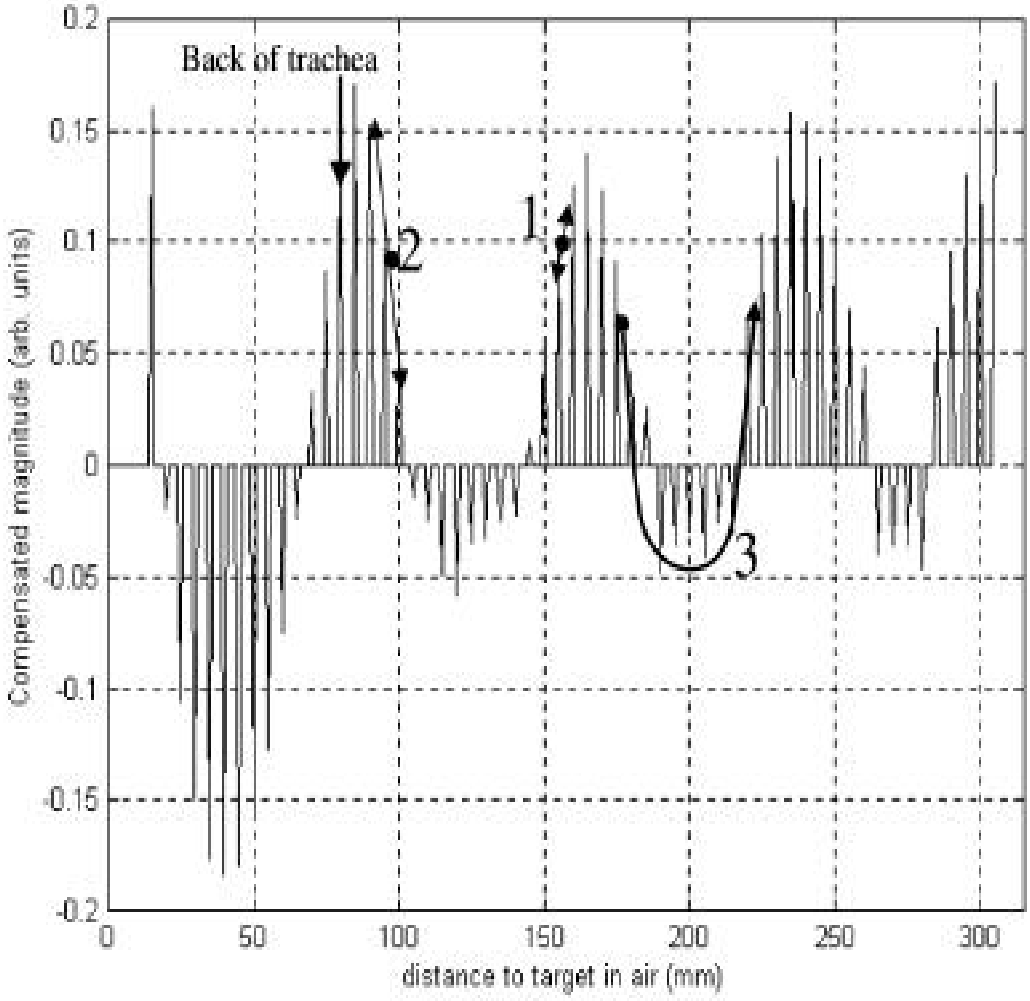

Figure 3. Homodyne sensor's AC-sensitivity curve as measured using a vibrating (~ 200 μm amplitude at 100 Hz) metal plate versus distance. Location 1, at 95 mm, indicates the sensor's small signal response of the rear tracheal wall motion (~ 50 μm) of the user. From this and other data, speech excitation information is obtained. Location 2, at 155 mm, indicates larger scale motion of a speech articulator, such as the 1.5-cm jaw-to-palate distance. Location 3, from 160 to 230 mm, illustrates how a > 7 cm motion of a user-interface device (e.g., a modulated resonant reflector) can be "tracked" and converted to relative distance traveled. The arrow showing the rear trachea wall is at 70 mm.

III. EXCITATION AND TRANSFER FUNCTIONS

A particularly interesting application of these sensors is to the measurement of windpipe wall and glottal structure motions, i.e., their amplitude versus time, as the vocal folds open and close. See the excitation signal insert in Fig. 4. The association of the EM sensor signal with vibrations of these particular tissue-air interfaces, was accomplished in several ways including moving the sensor relative to the neck to find the signal nulls and signal sign changes as the glottis opened and closed [3], using a laser velocimeter to measure rear tracheal wall motions [5], and performing extensive EM wave simulations [6]. These data enable an estimation of the voiced air pressure or air flow excitation function of human speech to be made [1,3]. The onset of excitation, causes an onset of a corresponding acoustic signal that reaches a typical acoustic microphone, at a time delayed by about 0.3 ms. The acoustic excitation generated at the glottis is transmitted, as well as being reflected and absorbed, as it travels up the vocal tract and out to the microphone. The reflecting and transmitting properties of the acoustic wave packet is a function of the wave frequencies and the tract's resonant and absorbing cavities. Well-established signal processing techniques such as ARMA (i.e., Auto-regressive, moving average) can be used to take advantage of a known (in our case, a real-time measured) excitation and acoustic output, and to then determine the filtering effects of the vocal tract. The data in Fig. 4

show an example of an excitation, the acoustic output, and the corresponding transfer function (TF) power spectrum taken as the sound /a/ (pronounced "ahh") was spoken. These ARMA methods enable more accurate "pole/zero" approximations of the transfer function to be used, than the most commonly used LPC methods. From the shape of the power spectrum of the transfer function, speech recognition algorithms can identify the sound being spoken as the phoneme /a/. Conversely, once excitation and transfer function information for an individually pronounced sound element is obtained, the sound can be reconstructed (i.e., synthesized). Similar experiments have been performed on stringed instruments, mechanical structures, and many other human speech sounds that have "source-filter" characteristics. The additional data obtained with excitation-measuring radar-like sensors can be used to more accurately characterize the sounds being produced, to identify the speaker, to synthesize their sounds, to cancel sounds, or to filter background noise [7].

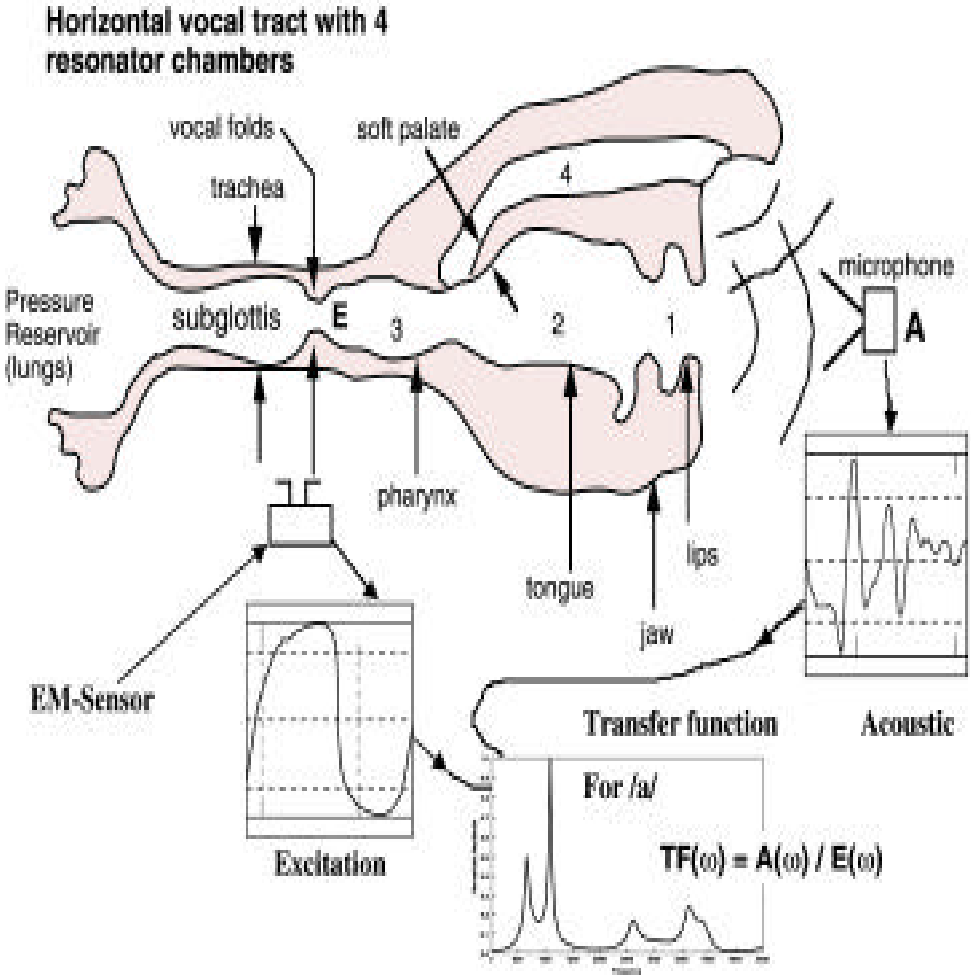

Figure 4. Midsagital plane of "linearized" human vocal system showing speech articulators, and EM sensor measurement locations. Major resonator chambers are numbered 1-4. The EM-sensor measures tissue motions as a consequence of sub-glottal pressure rise and fall, as the vocal folds open and close, and from the glottal structure directly. Signals illustrated include an excitation function, the output acoustic speech signal, and the corresponding transfer function (for the sound /a/) derived by deconvolution of the excitation from the output.

IV. HUMAN USE

The total radiated output of presently used EM sensors is 0.3 milliwatts, into 4π sterradians. When the sensor is placed close to the skin, about 1/2 of the output reaches the skin over an area of about 1.5 cm^2. Thus the average power on the tissue is < 0.1 mwatt/cm^2. This level is well below the U.S. standard for continuous user exposure to EM waves of 1.0 mW/cm^2 and is consistent with Swedish and Finnish standards of 0.1 mW/cm^2 for continuous exposure. In comparison, cellular telephones radiate about 0.5 W in similar frequency bands, leading to a flux of about 1 to 10 mW/cm^2 on the user's head and facial region, intermittently. The homodyne EM sensors described herein can be built into gener-

ally used appliances and communication devices, without any adverse effects to the user or to bystanders.

V. CONCLUSION

Low power EM radar-like sensors can measure generalized motions of animate and inanimate objects accurately, safely, and compactly. They are especially useful for non-contact measurements of motion parameters for those system components that are obscured by other dielectric materials, or that are extended in dimension, or when rapid (i.e., speech-of-light) measurements are needed. For applications near human users, their capacity to measure useful information, using low transmitted power, < 1 mW, is especially desirable. In particular, the homodyne EM sensor is proving to be very useful for measuring air-tissue interface motions of system excitation sources, as sound is produced. These enable detailed, real-time descriptions of the acoustic behavior of human speech systems, as well as from a wide variety of sound-producing mechanical systems. In addition, their large-scale motion-measurement properties are being applied to more conventional user interface devices such as "pointer" and "mouse-like" devices.

ACKNOWLEDGEMENTS

The authors would like to thank E. T. Rosenbury, and Drs. G. C. Burnett, and T. A. Gable for their assistance. Advice from Professors N. C. Luhmann and R. R. Freeman are also appreciated. The author thanks the U.S. Department of Energy and the National Science Foundation for their support. This work was performed in part under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

REFERENCES

[1] Holzrichter, J.F., Burnett, G.C., Ng, L.C., and Lea, W.A. "Speech Articulator Measurements Using Low Power EM Wave Sensor" *Journal Acoustic Society America* **103** (1) 622,1998. Also see the Website http://speech.llnl.gov/

[2] McEwan, T.E., U.S. Patent No. 5,345,471 (1994), U.S. Patent No. 5,361,070 (1994). U.S. Patent No. 5,573,012 (1996).

[3] Burnett, G.C., "The Physiological Basis of Glottal Electromagnetic Micropower Sensors (GEMS) and Their Use in Defining an Excitation Function for the Human Vocal Tract" Thesis University of California Davis, Jan. 15th, 1999, available on the Website mentioned in [1], and through University Microfilms, Ann Arbor, MI, document number 9925723.

[4] Skolnik, M. (1990). *Radar Handbook*, 2nd edition, McGraw-Hill, New York.

[5] Holzrichter, J.F, Kobler, J.B., Ng, L.C., Rosowski, J.J. Hillman, R.E., "Simultaneous laser doppler measurements of tracheal wall motions, EM sensor measurements of glottal structures, and sub-glottal pressure measurements" to be published.

[6] Burke, G.J., Champagne II, N.J., Holzrichter, J.F., Sharpe, R.M. "Simulation of EM wave scattering from human glottal structures" to be published.

[7] Ng, L.C.; Burnett, G.C.; Holzrichter, J.F.; and Gable, T.J. "Denoising of Human Speech Using Combined Acoustic and EM Sensor Signal Processing", Icassp-2000, Istanbul, Turkey, June 6, 2001.