UCRL-JC-134775
PREPRINT

Human Speech Articulator Measurements Using Low Power, 2 GHz Homodyne Sensors

J. F. Holzrichter
G. C. Burnett

This paper was prepared for submittal to the
24th International Conference on Infrared and Millimeter Waves
Monterey, California
September 5-10, 1999

June 29, 1999

Lawrence
Livermore
National
Laboratory

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Human Speech Articulator Measurements Using Low Power, 2 GHz Homodyne Sensors

J.F. Holzrichter* and G.C. Burnett

Lawrence Livermore National Laboratory and University of California, Davis

*Author correspondence: holzrichter1@llnl.gov

Abstract

Very low power, short-range microwave "radar-like" sensors can measure the motions and vibrations of internal human speech articulators as speech is produced. In these animate (and also in inanimate acoustic systems) microwave sensors can measure vibration information associated with excitation sources and other interfaces. These data, together with the corresponding acoustic data, enable the calculation of system transfer functions. This information appears to be useful for a surprisingly wide range of applications such as speech coding and recognition, speaker or object identification, speech and musical instrument synthesis, noise cancellation, and other applications.

Introduction

Recent studies using micro power radar-like sensors have shown that human speech articulator motions (see Holzrichter et al. [1]) and inanimate system vibrations can be measured in real time as acoustic sounds, such as speech, are produced. Initial work showed that very simple, non-spatially localized measurements provided information on a wide variety of generalized speech articulator motions—such as tissues associated with the glottal region, jaw, tongue, soft palate, lips and others. Similarly, characteristics of vibrating mechanical structures such as musical instrument strings, plastic or metal structures, and vibrating plates are easily measured. The primary mode of detection has been to measure EM sensor signals associated with changes in position of a system element versus time. However, internal filtering restricts measurement to those motions that take place in a determined time interval or a frequency band.

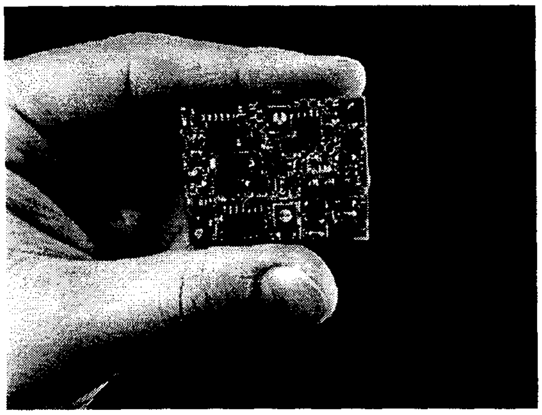

Figure 1. The microwave sensors transmit nominal 10 cycle pulse trains at 900 MHz or at 2.3 GHz, use a homodyne receiver mode, radiate < 0.5 mwatts of power, and use internal filtering. Two patch antennas transmit and receive 1.5 cm x 1 cm, are used.

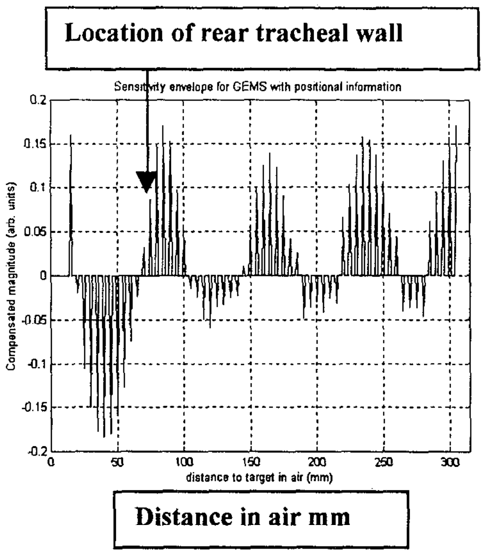

Figure 2. Homodyne sensor sensitivity curve as measured using a vibrating copper plate versus distance.

Homodyne Sensors

The homodyne field disturbance EM sensor works as an 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 [2]. The detection system and internal filters of 2 sensors used to date measure position change (as low as 2 micrometers) versus time, within frequency bands from 70 Hz to 7 kHz (sensor 1), and from 0.2 Hz to 10 Hz (sensor 2). The homodyne response function for a sensor designed to sense glottal tissue motions at 70-7kHz, is shown in Fig. 2. The response is from a 100 cm^2 Cu plate, vibrating at 200Hz with an amplitude of about 100 micrometers, that was moved for each data point, see Burnett [3]. When these data are used with non-unity dielectric systems, (e.g., human tissue with $\varepsilon = 50$) the corresponding effective wave path must be taken into account. The arrow in Fig. 2, at 70 mm, shows the effective position on the sensitivity curve of the rear windpipe wall of a subject. The association of the EM sensor signal with vibrations of this particular tissue-air interface was accomplished by moving the sensor relative to the neck (see Fig. 3) and noting the signal nulls and signal sign changes.

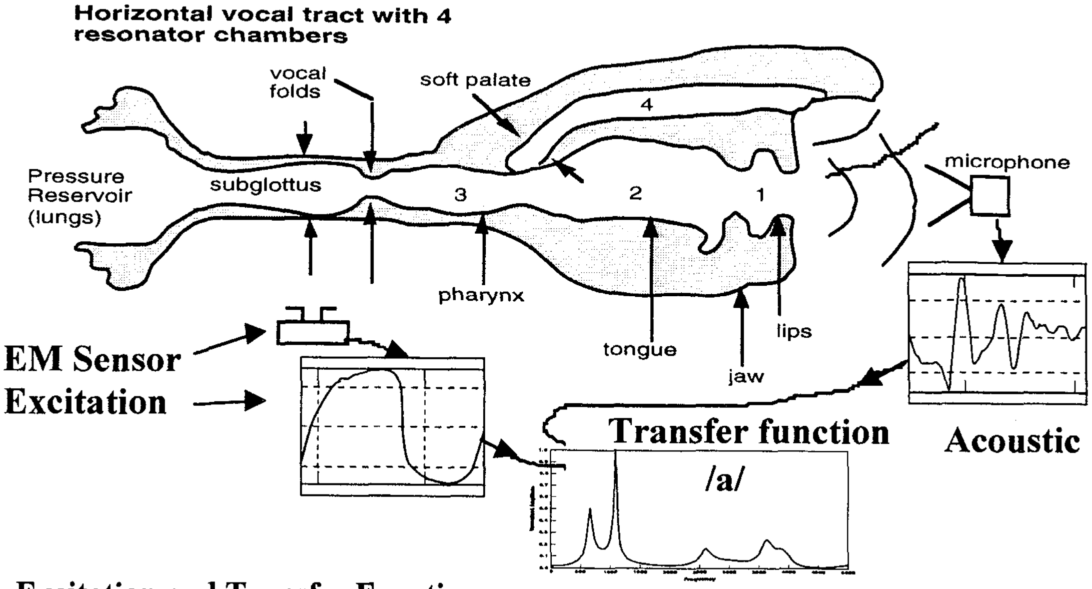

Figure 3. Midsagital plane of linearized human vocal system showing speech articulators, and EM sensor measurement locations. Typical signals include the excitation function, acoustic speech signal, and the transfer function for the sound /a/.

Excitation and Transfer Functions

The glottal region EM sensor measures wind pipe wall motion, i.e., amplitude versus time, in response to the subglottal pressure changes as the vocal folds open and close. See Fig. 3. These data enable an estimation of the voiced pressure excitation function of human speech to be made [3]. A short time after an excitation signal point (about 1.3 msec), the onset of the corresponding acoustic signal point is measured with a microphone. The excitation signal is reflected and absorbed, as a function of frequency, as it travels through the vocal tract and through air to the microphone. Using well-established signal processing techniques, the filtering effects of the resonators, between the glottal source and the microphone can be estimated. The data insert in Fig. 3 shows an example of the corresponding transfer function (TF) power spectrum taken as the sound /a/ (pronounced "ahh") was spoken. These methods enable more accurate "pole/zero" approximations of the TF to be used, compared to LPC methods. From the shape of the power spectrum of the transfer function, speech recognition algorithms can estimate the sound being spoken as the phoneme /a/. In addition, 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. Finally, other EM sensors can measure the changes in shapes as the resonator shapes change due to tongue, pharynx, lip, and other changes. These additional data can be used to more accurately estimate the sounds being produced.

Human Use

The total radiated output of presently used EM sensors is 0.3 milliwatts, into 4 pi 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 1W in similar frequency bands similar to those used with this the EM sensor herein, and thus about 1 to 10 mW/cm^2 reach the user's head and facial region, intermittently.

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 measuring motion parameters of system components that are obscured by other dielectric materials, that are extended in dimension, when rapid (i.e., speech-of-light) measurement is needed, and where low power sensors, near human users, are needed. 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 simple and detailed descriptions of the acoustic behavior of human speech systems and a wide variety of sound producing mechanical systems.

Acknowledgements

The authors would like to thank Drs. L. Ng, and S. Azevedo, Mr. T. Gable, E. T. Rosenbury, D. Poland, G. Dallum, P. Welch and C. Liang for their assistance. Advice from Professors N. Luhmann, and R. Freeman are appreciated. Also we thank the U.S. Department of Energy and the National Science Foundation for their support. Work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract 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 UC Davis, Jan. 15th, 1999, available on the Website mentioned in [1], and through University Microfilms.