Preprint
UCRL-JC-134183

Direct and Indirect Measures of Speech Articulator Motions Using Low Power EM Sensors

G. C. Burnett, J. F. Holzrichter, T. J. Gable, L. C. Ng

This article was submitted to XIVth International Congress of Phonetic Sciences, San Francisco, CA, August 1-7, 1999

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

May 12, 1999

Direct and Indirect Measures of Speech Articulator Motions Using Low Power EM Sensors

Burnett, G.C.; Holzrichter, J.F.; Gable, T.J.; and Ng, L.C.
Lawrence Livermore National Laboratory and University of California, Davis

ABSTRACT

Low power Electromagnetic (EM) Wave sensors can measure general properties of human speech articulator motions, as speech is produced. See Holzrichter, Burnett, Ng, and Lea, *J.Acoust.Soc.Am.* **103** (1) 622 (1998). Experiments have demonstrated extremely accurate pitch measurements (< 1 Hz per pitch cycle) and accurate onset of voiced speech. Recent measurements of pressure-induced tracheal motions enable very good spectral and amplitude estimates of a voiced excitation function. The use of the measured excitation functions and pitch synchronous processing enable the determination of each pitch cycle of an accurate transfer function and, indirectly, of the corresponding articulator motions. In addition, direct measurements have been made of EM wave reflections from articulator interfaces, including jaw, tongue, and palate, simultaneously with acoustic and glottal open/close signals. While several types of EM sensors are suitable for speech articulator measurements, the homodyne sensor has been found to provide good spatial and temporal resolution for several applications.

1. INTRODUCTION

Recent studies using micro power radar-like sensors have shown that speech articulators and related tissue motions can be measured in real time as acoustic speech is produced. Initial work by Holzrichter at al [1] showed that very simple, non spatially localized sensor measurements, can provide information on a wide variety of generalized articulator motions—such as tissues associated with the glottal region, jaw, tongue, soft palate, lips and others. The primary mode of detection has been to measure signal changes, occurring within a characteristic time interval (i.e., or frequency band) associated with a specific articulator. Thus the tissue interface motions associated with vocal fold opening and closing occur every 5 to 10 ms, in a frequency band around 100 to 200 Hz. The articulators that condition (i.e., transform) the excitation sounds from the glottis (and other sources) into speech sounds are in several locations, and move at rates of 1 Hz to 10s of Hz. With present EM sensors, the procedure for measuring specific articulator motions requires the speaker to pronounce only those phoneme sequences that involve a known single articulator motion. Two methods of determining such motions and their influences on speech have been pursued. They are 1) to measure the general articulator interfaces directly, and/or 2) to measure the transfer function, which is a consequence of the articulator motion; or to do both together.

To measure speech conditions involving multiple articulator gestures, the EM sensor should be capable of localizing the desired tissue motions of the target articulator, as speech is produced. Traditional radar range-gated sensors can, in principle, be used for precision articulator-interface location measurements, but they are not yet available. However, it has been found that a general homodyne EM sensor [1] can be used in an interferometric mode, where the oscillating sensitivity function versus distance can provide specific location information (see Fig. 2 below). Specific examples using time domain filtering to select articulator motions have been used by the authors and are described elsewhere, see refs [1], [2]. Examples of using spatial information from a homodyne sensor are discussed in detail in the recent thesis by Burnett [3]. Information obtained using two EM sensors, measuring two simultaneous articulators—the trachea and the soft palate—are illustrated in Fig. 1 and the data are shown in Fig. 3.

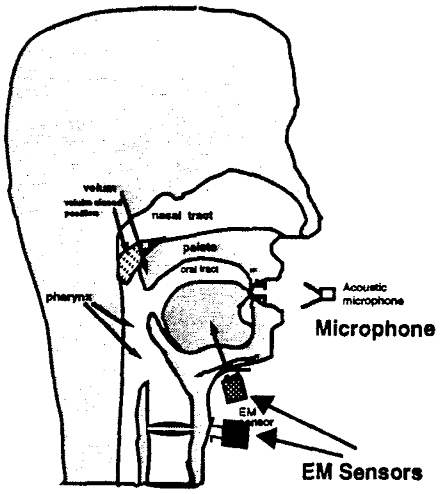

Figure 1: Midsagittal view of speech articulators, showing two EM sensor positions, and a microphone location.

2. HOMODYNE SENSORS

The homodyne field disturbance EM sensor works as an interferometer by measuring the reflection of a transmitted wave against a local (phase reference) wave. As the reflecting surface moves, the phase of the reflected wave varies with respect to the stationary local wave; and a signal associated with this change is detected by a mixer and filter combination and is shown below in Fig. 2.

The sensitivity function illustrated in Figure 2 demonstrates the change in amplitude of the mixed signal that occurs in a homodyne receiver. As the sensor moves away from a reflecting interface, the amplitude of the return signal changes due to the change in phase of the reflected wave. This calibration was done in air, so any transmission that occurs through a non-unity dielectric must be compensated by the square root of the dielectric constant ε. After compensation, the equivalent distance in air and the sensitivity curve are used to calculate the expected magnitude of the sensor signal. Conversely, if the magnitude of the signal is known at several points (the simplest is where the magnitude is zero), and the equivalent distance to the reflecting surface is not known, the equivalent distance from the sensor to the reflecting surface can be calculated. The in-tissue distance can then be estimated using the dielectric constants of the tissue layers at the frequency of interest (around 2 GHz). In this manner, the submillimeter pressure-induced motions of the rear tracheal wall, as the vocal folds opened and closed, were obtained by Burnett [3]. He used this sensor and noted that the rear trachea wall behaved as if it were 81 ± 3 mm in air. This distance in tissue is equivalent to (air path) / $(\varepsilon)^{1/}$= 3.5 cm, which is the location of his rear tracheal wall, as obtained from CT images. Using this tracheal wall as a pressure sensor, a voiced excitation function is obtained. With it, excellent transfer functions are calculated for each sound of human speech. See examples in [1] and in the lectures available at the referenced Website.

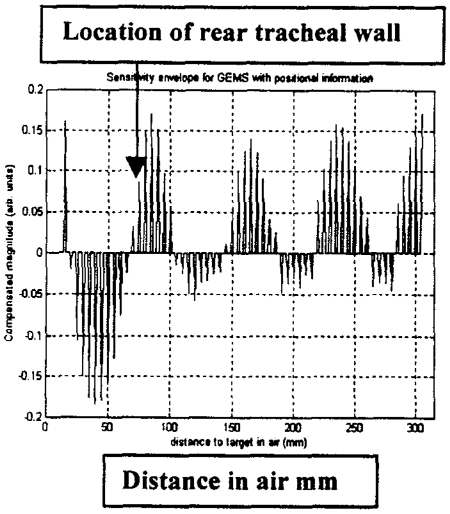

Figure 2: Homodyne sensor "interferometer-like" sensitivity function. Arrow shows sensitivity curve for tracheal tissue motion.

3. COMBINED SENSORS

In order to directly measure several speech articulators at once; an experiment was set up as shown in Fig. 1. Two EM homodyne sensors were used, and a single articulator, the velic port, was exercised (to the degree possible by the authors) to affect acoustic speech signal changes. This approach is illustrated by using one sensor to measure the motions of the velum (and possibly some tongue hump motion) as the sequence of speech /a/ /ng/ /a/ was produced, at the same time the tracheal tissue motions were measured and the acoustic signal was recorded. This enabled a voiced excitation function to be estimated and a transfer function to be obtained for every two glottal cycles. The transfer functions, shown at the bottom of Fig. 3 on an expanded scale with 16 ms between frames, is positioned to correspond to the spectrogram, and to the measured velum motions.

The data in Fig. 3 shows the general gesture of the soft palate, but it is offset in time from the appearance (or disappearance) of the nasal spectral properties. Since it is difficult to use the EM sensor to measure the actual distances of palate movement associated with the onset of airflow, or with the production of an air seal, a simultaneous transfer function offers a very precise method of such measurements.

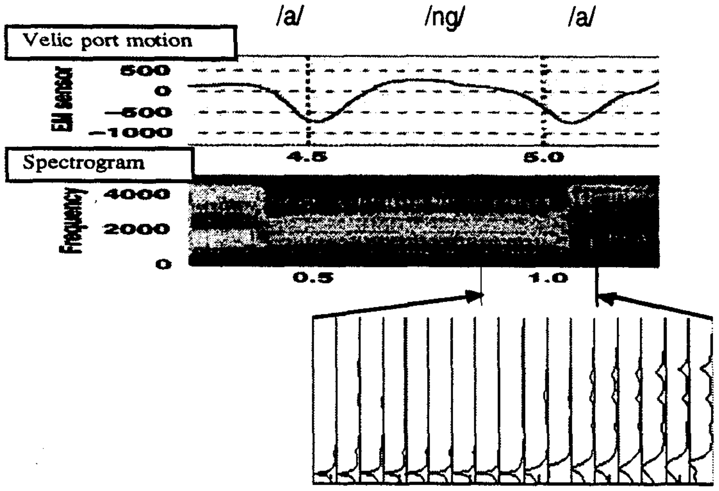

Figure. 3: An EM sensor signal showing velum motion, a simultaneous acoustic spectrogram, and sequential 16 ms frames of the corresponding vocal tract transfer function. The speaker produced a sequence of phonemes: /a/ /ng / /a/

4. CONCLUSION

Low power EM radar-like sensors can measure generalized motions of articulators quite accurately. Their capacity to distinguish between different articulators, operating at the same time, depends upon the spatial and temporal differences of the measured articulator interfaces

The homodyne EM sensor is proving to be quite flexible in that it enables very small articulator interface motions to be detected as well as providing important location information. These have lead to an accurate estimate of a voice excitation function, which enables the calculation of a transfer function. The combination of transfer function information, with direct articulator measurements can provide otherwise difficult to obtain information on the precise timing of several articulators working together to produce speech.

ACKNOWLEDGMENTS

The authors would like to thank Prof. John Ohala for his continued encouragement of this work and Dr. Rebecca Leonard (UC Davis Medical Center) for her assistance. Also we thank the US Department of Energy and the National Science Foundation for their support.

This work was performed under the auspices of the U.S. DOE by LLNL 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] Burnett, G.C.; Gable, T.J.; Holzrichter, J.F.; Ng, L.C. "Accurate and noise-robust pitch extraction using low power electromagnetic sensors" submitted for publication 1998

[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].