INSTITUTE OF PHYSICS PUBLISHING MEASUREMENT SCIENCE AND TECHNOLOGY

Meas. Sci. Technol. **15** (2004) 1291–1302 PII: S0957-0233(04)73072-6

A novel non-acoustic voiced speech sensor

D R Brown III, R Ludwig, A Pelteku, G Bogdanov and K Keenaghan

Worcester Polytechnic Institute, Electrical and Computer Engineering Department, 100 Institute Road, Worcester, MA 01609, USA

E-mail: drb@wpi.edu

Received 5 December 2003, in final form 28 April 2004
Published 16 June 2004
Online at stacks.iop.org/MST/15/1291
doi:10.1088/0957-0233/15/7/010

Abstract

Non-acoustic speech sensors have a long history of clinical applications but have only recently been applied to the problem of measuring speech signals in the presence of strong background noise. These sensors typically provide measurements of one or more aspects of the speech production process, such as glottal activity, as a proxy for the actual speech and tend to be highly immune to acoustic noise. In this paper, a new non-acoustic speech sensor based on a tuned electromagnetic resonator collar is proposed. The collar is designed to be worn around the talker's neck and is sensitive to small changes in the dielectric properties of the glottis as well as subglottal and supraglottal systems that result from voiced speech. Unlike the majority of previously developed non-acoustic speech sensors, the proposed sensor does not require skin contact or precise alignment to effectively measure glottal activity. This paper develops the sensor concept and provides analytical, simulated and experimental results that demonstrate the potential of the new speech sensor.

Keywords: voiced speech, pitch estimation, capacitive sensors, electroglottogram

1. Introduction

A promising new approach to the problem of measuring speech signals in the presence of strong background noise is the use of non-acoustic speech sensors. A non-acoustic speech sensor measures aspects of the speech generation process as a proxy for the actual acoustic speech signal and tends to be highly immune to acoustic noise. These sensors have been historically used almost exclusively in clinical environments for applications such as pitch determination, but recent research has begun to consider their potential for improving speech quality in high-noise environments. Non-acoustic speech sensors typically do not measure enough information about the speech generation process to replace microphone sensors; instead, these sensors are commonly used in conjunction with a microphone and additional signal processing in order to augment the acoustic speech signal and to improve the resulting speech quality. In some cases, the gains in speech quality can be remarkable (Ng *et al* 2000). In addition to improving speech quality, other useful applications for this technology include improved talker authentication/identification (Campbell *et al* 2003) and very low bit-rate voice coding (Brady *et al* 2004).

One example of a commonly used non-acoustic speech sensor is the electroglottogram (EGG) (Boves and Cranen 1982, Baken 1992). The EGG consists primarily of a pair of electrodes placed in contact with the talker's neck and a signal generator used to establish a small sinusoidal voltage across the electrodes. Glottal activity is detected by measuring the change in electrical impedance across the throat that occurs during voiced speech. Other examples of non-acoustic speech sensors include ultrasonic or photoelectric sensors (Hess 1983), bone- or skin-conduction accelerometers (Scanlon 1998) and low-power radar-based sensors (Holzrichter *et al* 1998, Burnett 1999).

Radar-based non-acoustic speech sensors have been the focus of recent research in this area. These sensors employ a high frequency (e.g. 2 GHz) transmitter/receiver pair and operate on the principle of round-trip-travel-time, where the position of a reflecting object (e.g. a tracheal wall) can be inferred from the propagation time between the transmission and reception of a radar pulse. Laboratory experiments

with low-power radar-based sensors have shown that these sensors may be effective at measuring the position of vocal articulators as well as glottal activity during voiced speech segments. The measured data have been used to form accurate pitch and vocal tract transfer function estimates (Burnett *et al* 1999a) as well as to provide a method for effective noise removal from acoustically noisy speech signals (Ng *et al* 2000).

While low-power radar-based speech sensors show promise in some applications, these sensors typically require that the transmit/receive antenna be placed in direct contact with the talker's skin in a suitable location. Radar's inherent reliance on accurately measuring the signal's round-trip-travel-time from a reflecting object implies that radar-based speech sensors may be sensitive to transmitter/receiver alignment and positioning. Moreover, complicated reflective environments may obscure the desired target of a radar-based vocal function sensor and lead to ambiguity as to what is actually being measured (Burnett *et al* 1999b).

In this paper, we propose a new non-acoustic speech sensor based primarily on a tuned electromagnetic resonator collar (TERC) by leveraging technology recently developed for use in magnetic resonance imaging (Bogdanov and Ludwig 2002, Ludwig *et al* 2004). We provide analytical, simulated and experimental results that demonstrate the potential of the TERC sensor. Like the majority of non-acoustic speech sensors, including the radar-based sensors and the EGG, the TERC sensor is designed to measure glottal activity during voiced speech. The TERC sensor, however, resolves several shortcomings in the present technology. Unlike most radar-based sensors, the TERC sensor does not require direct skin contact and does not directly measure the position of vocal articulators or use round-trip-travel-time measurements. Instead, the TERC sensor measures the *integrated behaviour* of a cross section of neck tissue that includes the glottis as well as subglottal and supraglottal systems. The TERC sensor does not require precise positioning or alignment and is inherently robust to the complicated reflective environment of the neck. Also, unlike the EGG, the TERC sensor does not apply any voltage or current directly to the talker.

The paper is organized as follows: section 2 describes the basic concept and the principle of operation of the proposed non-acoustic TERC speech sensor. Section 3 presents two analysis and simulation methods that facilitate selection and optimization of the design parameters of the TERC speech sensor. This section also contains a specific sensor design example and applies the described analysis and simulation methods to obtain performance predictions for this particular case. Section 4 presents experimental results, obtained by constructing a physical prototype of the TERC sensor, which verify the analytical results and also demonstrate the basic principle of operation. Finally, section 5 presents our conclusions and outlines potential future research directions.

2. Tuned electromagnetic resonator collar speech sensor

This section describes the basic speech production mechanism and physiological properties that are relevant to the operation of the TERC speech sensor. We then present the TERC sensor concept and discuss its principle of operation. This is followed by a discussion of typical TERC sensor design parameters.

2.1. Voiced speech production and physiological properties

Like the majority of non-acoustic speech sensors, the TERC sensor is intended to measure voiced speech. The source of acoustic energy for all voiced speech is the glottal cycle (Stevens 1997, Titze 1980), which can be summarized as follows: the vocal folds are pulled into a closed state by the laryngeal muscles; expiratory air flow from the lungs causes air pressure to build in the area behind the vocal folds; the vocal folds are forced open for a brief period of time when the air pressure exceeds the retaining force of the laryngeal muscles; the vocal folds recoil to the closed state after a small puff of air escapes. During voiced speech, the glottal cycle is repeated at a frequency equal to the fundamental frequency of the acoustic waveform produced at the output of the vocal tract. This property of speech production is exploited by almost all non-acoustic speech sensors. For example, the EGG uses the glottal cycle and the fact that movement of the vocal folds leads to variations in the transverse electrical impedance of the neck in order to measure the glottal activity induced by voiced speech.

The physiological property of the glottal cycle relevant to the TERC speech sensor, and the property that makes this sensor unique among non-acoustic speech sensors, is that the relative permittivity of most body tissue consistent with the location of the glottis is on the order of 50–200 times that of air for frequencies in the range of 10 MHz–200 MHz (Bronzino 1995, chapter 90). Intuitively, this implies that when the glottis is open and an air cavity is present in the larynx, the composite relative permittivity of a cross section of the neck through the larynx is lower than when the glottis is closed. During voiced segments of speech when the glottal cycle is present, these facts imply that the composite relative permittivity of a cross section of the neck through the larynx oscillates at a frequency equal to the fundamental frequency of the acoustic waveform.

2.2. Sensor concept and principle of operation

The physiological properties discussed in the previous section motivate the development of a sensor that is able to measure changes to the relative permittivity of the larynx as a proxy for measuring movement of the glottis (as well as movement of tissue in the subglottal and supraglottal systems). A common method of measuring the relative permittivity of a particular material is to establish an electric field through the material by constructing a capacitor (or an array of capacitors) and to measure the resulting capacitance. For example, consider a simple parallel plate capacitor with time-varying dielectric permittivity. The time-varying capacitance can be expressed as

$$C(t) = \frac{\epsilon(t)A}{d} \tag{1}$$

where $\epsilon(t)$ denotes the time-varying permittivity of the material between the capacitor plates, A denotes the constant area of the plates and d denotes the constant separation between the plates. Since capacitance is proportional to

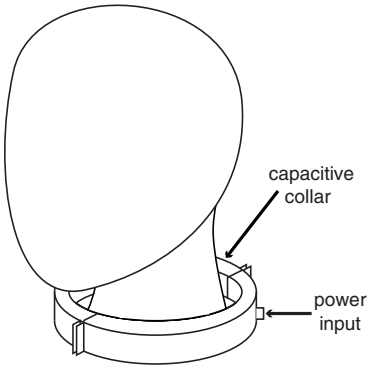

Figure 1. TERC speech sensor concept.

permittivity, measuring changes in the capacitance provides a straightforward method for measuring changes in the permittivity.

With the basic principles thus established, the TERC speech sensor concept is shown in figure 1. One or more capacitors are constructed around the neck tissue by placing two or more conductive plates on a collar around the talker's neck. The collar is not required to be in contact with the neck, but may be worn this way for convenience. Insulation is placed between any exposed conductive plates and the talker's neck to prevent skin contact and undesirable electrical conduction.

A fundamental difficulty in this approach is the problem of constructing the sensor so as to be sensitive to small changes in the relative permittivity of the neck. The accuracy of typical capacitance measurement methods may not be sufficient to clearly distinguish the glottal cycle. Our approach to this problem is to monitor the complex impedance of the sensor. The sensor input impedance, denoted as $Z_{\text{in}}(f)$, is a function of the sensor design parameters, the frequency f and the effective relative permittivity of the material in the sensor interior (which is itself a function of the frequency f). The glottal cycle can be expected to cause only small perturbations to $Z_{\text{in}}(f)$ over most or all of the applicable frequency range. Nevertheless, if the sensor is driven by an RF source matched to the characteristic impedance Z_o of the coaxial cable, the reflection coefficient of the sensor can be expressed as

$$S_{11}(f) = \frac{Z_{\text{in}}(f) - Z_o}{Z_{\text{in}}(f) + Z_o}. \tag{2}$$

The magnitude shift of this parameter is easily measured by a standard network analyser.

We define a 'resonant' frequency of the sensor as a frequency point f_r where $|S_{11}(f_r)| \approx 0$ (or, equivalently, $|Z_{\text{in}}(f_r) - Z_o| \approx 0$). If the sensor is designed to have one or more resonant frequencies and is driven by the RF source at a resonant frequency ($f_{\text{RF}} = f_r$), small changes in $Z_{\text{in}}(f_{\text{RF}})$ can be measured by observing relatively large changes in the reflection coefficient at this frequency. In other words, the sensor is designed such that small changes in the permittivity of the material in the interior of the sensor are manifested as relatively large changes (depending on the quality factor of the resonator) in the reflection coefficient of the sensor. This sensitivity enhancing behaviour is summarized in figure 2.

Unfortunately, it can be difficult to design the sensor to always maintain one or more deep resonances at consistent frequencies for a variety of talkers. A practical solution to this problem is the addition of a lumped tuning and matching circuit (constructed of tunable inductors and capacitors) placed between the sensor and the RF source. These lumped elements are tuned with the sensor placed on the talker to locate the resonances in the desired frequency range and to optimize the quality factors of the resonances. Once the tuning is complete, the sensor is driven by an RF source at low power (typically 0 dBm or less) at a frequency close to one of the resonances. Voiced speech is then measured in the perturbations to the reflection coefficient seen by the RF source.

2.3. Typical design parameters of the TERC speech sensor

Our basic approach to the physical design of the capacitive collar leverages recent work in the area of coupled microstrip line transverse electromagnetic (TEM) resonators (Bogdanov and Ludwig 2002, Ludwig *et al* 2004). This technology was originally developed for high-field magnetic resonance imaging, but is easily adapted to the general design of a collar-shaped capacitive sensor. Figure 3 shows a typical TERC sensor where two or more 'microstrip lines' are placed

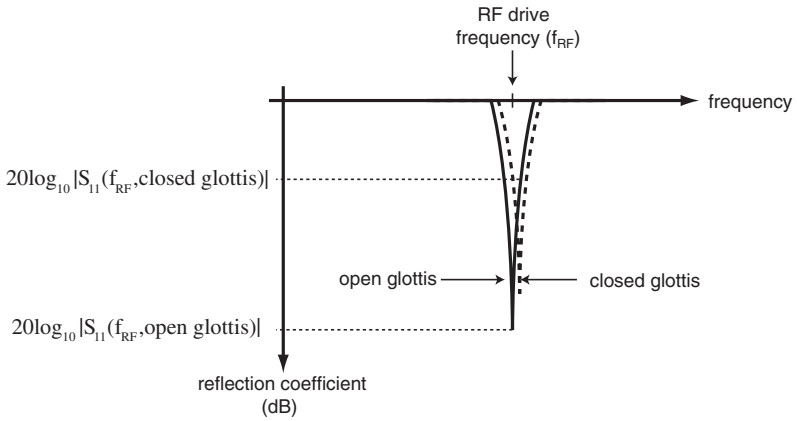

Figure 2. Illustration of the sensitivity enhancement principle of the TERC speech sensor where small shifts in the sensor resonant frequency result in relatively large changes in the sensor reflection coefficient.

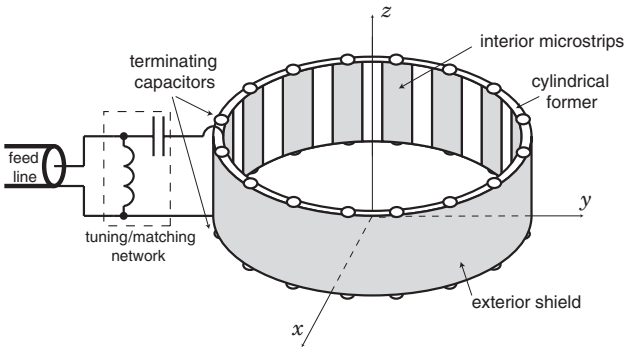

Figure 3. Generic TERC speech sensor construction example. An arbitrarily selected interior microstrip line serves as the drive port that is connected to the feed line through the matching/tuning network.

on the interior of a non-conductive cylindrical former (constructed of thermoplastic material such as LEXAN[1]) and a 'shield' is placed on the exterior of the former. The interior microstrip lines are connected to the exterior shield by terminating capacitors located on the top and bottom edges of the cylindrical former. The RF source is connected to the collar, through the tuning and matching network, by attaching the signal lead to one of the microstrip lines and the ground lead to the collar shield. To make this system useful as a speech sensor, the system is either split (as shown in figure 1) or constructed on an arc rather than a full cylinder.

Under the basic design constraints of figure 3, there are still several design parameters that influence the number of resonant frequencies, the location and depth of the resonant frequencies, the field penetration and the overall sensitivity and bandwidth of the TERC speech sensor. These parameters include:

(i) The inner and outer radii of the cylindrical former.
(ii) The height of the cylindrical former.
(iii) The number of interior microstrip lines.
(iv) The width of the interior microstrip lines.
(v) The spacing between neighbouring interior microstrip lines.
(vi) The value of the terminating capacitors.

While it is impossible to give an in-depth treatment of all these parameters here, the following section describes two analytical methods that we have successfully used to evaluate the impact of these parameters for specific sensor designs. These methods are general in the sense that they can be applied to both the design and performance evaluation of TERC sensors constructed according to the basic guidelines shown in figure 3.

3. Design and performance analysis

This section describes two methods for designing and evaluating the performance characteristics of the TERC speech sensor. The first method is the multi-transmission line (MTL) model. The MTL method offers a computationally efficient method for rapid design and evaluation of the reflection coefficient of the TERC sensor and has been extensively used to predict the behaviour of resonating coils in high-field magnetic resonance imaging (Bogdanov and Ludwig 2002). While the MTL method is quite general and can quickly approximate the overall behaviour of the sensor, it can only be used within the constraints of the transverse electromagnetic theory excluding radiation effects and conductive loss mechanisms in the glottal load models. The second and more elaborate design/evaluation method presented in this section is the finite element (FE) analysis approach that allows for more complicated glottal load models. The FE analysis method generally provides accurate results but also requires significantly increased computational resources in terms of memory and cycle time. Finally, this section concludes with a specific TERC sensor design example, the performance of which is evaluated with the MTL and FE approaches.

3.1. Multi-transmission line model rapid design/evaluation

The electromagnetic field distributions of a cylindrical (but not necessarily cylindrically symmetric) electromagnetic resonator can be conveniently approximated with a MTL model. In particular, this model can serve as a rapid design tool for predicting the $S_{11}(f)$ magnitude response of the TERC sensor as well as changes to the $S_{11}(f)$ magnitude response that occur as a result of changes in a simple dielectric load in the interior of the sensor. Although the MTL method was originally developed in Bogdanov and Ludwig (2002) for applications in magnetic resonance imaging, it is important to review the key principles here in order to describe how the MTL method can be applied to the design of the TERC speech sensor.

The formulation in the frequency domain employs a set of generalized transmission line equations that are written in matrix form involving spatially varying voltage and current waves (Paul 1994)

$$\frac{\mathrm{d}}{\mathrm{d}z}\begin{bmatrix}\boldsymbol{v}(z)\\ \boldsymbol{i}(z)\end{bmatrix}=\begin{bmatrix}\mathbf{0} & -\boldsymbol{Z}\\ -\boldsymbol{Y} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{v}(z)\\ \boldsymbol{i}(z)\end{bmatrix} \tag{3}$$

where $\boldsymbol{v}(z)$ and $\boldsymbol{i}(z)$ are column vectors representing the voltage and current distributions at discrete spatial locations along the longitudinal axis of the sensor (aligned with the z-axis shown in figure 3) and $\boldsymbol{Z} = \boldsymbol{R} + \mathrm{j}\omega\boldsymbol{L}$ and $\boldsymbol{Y} = \boldsymbol{G} + \mathrm{j}\omega\boldsymbol{C}$ are the per-unit-length impedance and admittance matrices, respectively. These matrices characterize the physical behaviour of the transmission line configuration as a function of angular frequency $\omega = 2\pi f$ and per-unit-length resistance $\boldsymbol{R}$, conductance $\boldsymbol{G}$, inductance $\boldsymbol{L}$ and capacitance $\boldsymbol{C}$ matrices. As discussed in Bogdanov and Ludwig (2002), a standard electrostatic boundary element technique is invoked to compute these matrices for an xy-plane cross sectional slice of the sensor shown in figure 3. Following basic linear, first order differential matrix equation theory, (3) has the general solution

$$\begin{bmatrix}\boldsymbol{v}(z)\\ \boldsymbol{i}(z)\end{bmatrix}=\underbrace{\begin{bmatrix}\boldsymbol{\Phi}_{11}(z) & \boldsymbol{\Phi}_{12}(z)\\ \boldsymbol{\Phi}_{21}(z) & \boldsymbol{\Phi}_{22}(z)\end{bmatrix}}_{:=\boldsymbol{\Phi}(z)}\begin{bmatrix}\boldsymbol{v}(0)\\ \boldsymbol{i}(0)\end{bmatrix} \tag{4}$$

where $\boldsymbol{\Phi}(z)$ is the so-called chain parameter matrix (Bogdanov and Ludwig 2002) defined as

$$\boldsymbol{\Phi}(z) = \mathrm{e}^{z\boldsymbol{A}}; \qquad \boldsymbol{A} = \begin{bmatrix}\mathbf{0} & -\boldsymbol{Z}\\ -\boldsymbol{Y} & \mathbf{0}\end{bmatrix}. \tag{5}$$

[1] LEXAN is a trademark of the General Electric Company.

The unknown constant vectors $\boldsymbol{i}(0)$ and $\boldsymbol{v}(0)$ in (4) can be found from the terminating (boundary) conditions at the source and load sides. Assuming that the load side contains no sources, the sensor termination conditions at $z = 0$ (the bottom edge of the cylinder) and $z = L$ (the top edge of the cylinder) are found from generalized Thévenin equivalent circuit expressions

$$\begin{aligned} \boldsymbol{v}(0) &= \boldsymbol{Z}_{\text{in}}\boldsymbol{i}(0) \\ \boldsymbol{v}(L) &= \boldsymbol{Z}_L\boldsymbol{i}(L) \end{aligned} \tag{6}$$

where the unknown $\boldsymbol{Z}_{\text{in}}$ is the input impedance matrix of the MTL system of length L terminated by a load network $\boldsymbol{Z}_L$. Since $\boldsymbol{Z}_L$ typically incorporates lumped elements such as capacitors, the MTL formulation can directly combine lumped tuning elements with the distributed transmission line microstrip lines. Tuning these capacitors results in a shift of the resonance frequency spectrum.

Once (4) and (6) are determined, the input impedance matrix $\boldsymbol{Z}_{\text{in}}$ is computed and related to the input reflection coefficient via

$$\boldsymbol{\Gamma}_{\text{in}} = [\boldsymbol{Z}_{\text{in}} - \boldsymbol{Z}_0][\boldsymbol{Z}_{\text{in}} + \boldsymbol{Z}_0]^{-1} \tag{7}$$

where $\boldsymbol{\Gamma}_{\text{in}}$ and $\boldsymbol{Z}_0$ are diagonal matrices representing the individual reflection coefficients and characteristic line impedances (typically 50 Ω) for each of the microstrip lines. As shown in figure 3, only one microstrip line is driven by the RF source and the remaining microstrips are terminated by capacitors. The resulting reflection coefficient of this one-port system is then given as $S_{11} = [\boldsymbol{\Gamma}_{\text{in}}]_{n,n}$ where n is the index of the driven microstrip line.

3.2. Finite element frequency domain design/evaluation

While the MTL model serves as an indispensable rapid design/evaluation tool that can quickly predict the behaviour of a given sensor design as well as assist in the specification of distributed and lumped components, it suffers from the drawback that it cannot take into account radiation effects and lossy biological loadings such as the larynx with the embedded glottis. For this reason, a full-wave three-dimensional finite element (FE) formulation is adopted that allows the modelling of radiation effects in dielectric media of complicated shapes. Unlike the MTL model, which treats Maxwell's equations within the constraints of a transverse electromagnetic theory, the FE formulation does not suffer from similar restrictions since it is governed by the double curl vector wave equation for the electric field

$$\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \boldsymbol{E} = \omega^2\mu\epsilon\boldsymbol{E} - \mathrm{j}\omega\mu\sigma\boldsymbol{E} \tag{8}$$

and an identical equation for the magnetic field. The FE solution ensures continuity of the electric field, while the magnetic field can jump by an impressed surface current density at interfaces between different materials

At surfaces where the FE mesh is terminated, additional boundary conditions should be imposed that describe either an electric wall or a magnetic wall. For example, copper strips are applied as a perfect electric boundary in order to limit mesh size and memory usage. In the case of an open region, however, the FE mesh must be terminated into a radiation boundary which attempts to absorb all radiating fields and eliminate undesired reflections from the truncated mesh. We have adopted the idea of a perfectly matched layer (PML), as initially proposed for solving finite difference time domain problems (Berenger 1994) and later adopted for FE formulations (Sacks *et al* 1995). This technique describes an idealized medium that approximates a reflectionless surface for all single frequency incident waves between the solution domain and the PML region. Consistent with Maxwell's theory, the divergence equation and Ampère's law for such a medium can be stated as

$$\begin{aligned} \boldsymbol{\nabla} \cdot [\epsilon_{PML}]\boldsymbol{E} &= 0 \\ \boldsymbol{\nabla} \times \boldsymbol{H} &= \mathrm{j}\omega[\epsilon_{PML}]\boldsymbol{E} + \boldsymbol{J} \end{aligned} \tag{9}$$

where we have used the diagonally anisotropic material tensor

$$[\epsilon_{PML}] = \epsilon_0 \begin{bmatrix} \epsilon_x - \mathrm{j}\left(\sigma_E^x/\omega\right) & 0 & 0 \\ 0 & \epsilon_y - \mathrm{j}\left(\sigma_E^y/\omega\right) & 0 \\ 0 & 0 & \epsilon_z - \mathrm{j}\left(\sigma_E^z/\omega\right) \end{bmatrix}. \tag{10}$$

Here σ_E^x, σ_E^y and σ_E^z are the electric conductivities in the x, y and z directions, respectively. Our choice for implementing the PML method is to enclose the sensor within a sufficiently large air box and then add the perfectly matched layers with appropriate tensor values.

Over the surface of the sensor, impedance boundaries are enforced such that

$$\boldsymbol{E}_t = -Z_s\hat{\boldsymbol{n}} \times \boldsymbol{H} \tag{11}$$

and

$$\boldsymbol{H}_t = \frac{1}{Z_s}\hat{\boldsymbol{n}} \times \boldsymbol{E}. \tag{12}$$

Here Z_s is the surface wave impedance and $\hat{\boldsymbol{n}}$ is the outward normal to the surface of the solution space. Similar to (6) in the MTL approach, equations (11) and (12) provide the theoretical framework for handling lumped impedance elements, such as the terminating capacitors depicted in figure 3. Polylines are used to discretize surfaces connecting the interior microstrips to the exterior shield; these surfaces emulate the lumped terminating capacitors shown in figure 3.

3.3. Sensor design example and numerical predictions

This section provides a specific example of a TERC speech sensor design and applies the MTL and FE methods developed in the prior sections to select terminating capacitors and evaluate the sensitivity of the proposed sensor to changes in the glottal state. The results obtained in this section are compared to experimental results in section 4.

The example sensor design is similar to the basic construction shown in figure 3 with specific parameters given in table 1. Note that these parameters correspond to a design where the microstrips and shield cover an arc of 120° on the cylindrical former. The remaining 240° of the cylindrical former are not used. The prototype configuration specifies six interior microstrip lines that result in six resonant modes. The choice of six elements represents a heuristic compromise between design complexity and deploying a sufficient number of elements near the glottal region. The prototype configuration also specifies a target resonant frequency of 200 MHz. This resonant frequency

Table 1. TERC speech sensor prototype design parameters.

Parameter	Value
LEXAN former inner radius	69.8 mm
LEXAN former outer radius	76.2 mm
LEXAN former height	38.1 mm
Number of interior microstrip lines	6
Width of interior microstrip lines (same for all)	21.21 mm
Spacing between interior microstrip lines (same for all)	3.175 mm
Total angle spanned by shield and interior microstrip lines	120°
Value of terminating capacitors	MTL = 95 pF, FE = 108 pF

Table 2. Dielectric properties of relevant tissues at 200 MHz. These values were computed using the interactive application written by the Italian National Research Council, Institute for Applied Physics at http://niremf.iroe.fi.cnr.it/tissprop/.

Body tissue	Relative permittivity	Conductivity
Skin	55.716	0.582 29
Muscle	60.228	0.743 07
Cartilage	49.161	0.517 51
Blood	68.474	1.280 2
Blood vessel	51.088	0.508 61
Bone	18.2	0.13
Spinal cord	39.70	0.385 02

was selected since the wavelength in muscle ($\epsilon_r \approx 60$) is $\lambda = c/(f\sqrt{\epsilon_r}) \approx 19$ cm, which is only slightly larger than the diameter of a human neck while maintaining an acceptable penetration depth (skin depth) of $\delta = 1/\sqrt{\pi f \mu_0 \sigma} \approx 6$ cm. The detailed FE simulations presented later in this section show that certain modes tend to exhibit stronger field penetration into the glottis which corresponds to greater sensitivity to changes in the glottal state.

Given this configuration, the MTL method was initially applied to the problem of computing values for the terminating capacitors such that the first resonant mode would be placed at 200 MHz. The appropriate value was computed to be 95 pF through iterative application of the MTL equations. MTL analysis was then used to estimate the overall $S_{11}(f)$ magnitude response of the sensor in two states: glottis open and glottis closed. Ideally, an anatomically accurate tomographical slice of the larynx would be used to model the load but, since the MTL model is constrained to geometrically simple loads with nonconductive losses, only approximate biological models of these states can be employed. The load model for the closed glottis state is a solid cylinder of muscle with radius of 65 mm and a height of 38.1 mm centred in the interior of the collar. The relative permittivity of muscle at 200 MHz is given in table 2. The load model for the open glottis state is the same cylinder of muscle except that a 12 mm radius cylinder of muscle offset 50 mm from the centre of the neck is removed and replaced by air. Figure 4 shows the overall predicted $S_{11}(f)$ magnitude response in the open and closed glottis states and figures 5 and 6 show detailed plots of the second and third resonant modes.

The MTL analysis results in figure 4 clearly show the expected six characteristic resonances, consistent with the fact that the sensor contains six microstrip elements. While

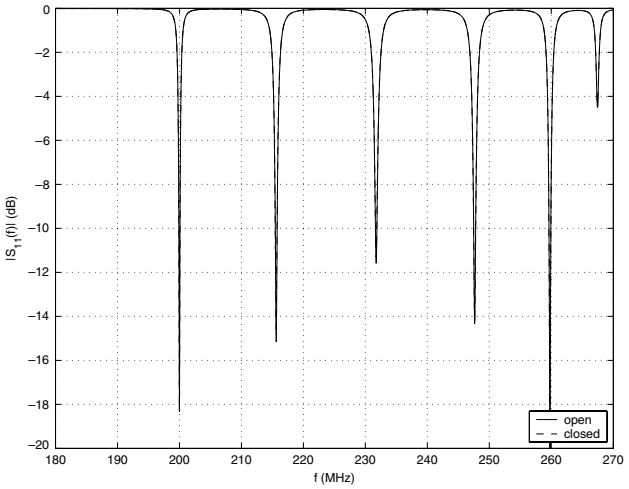

Figure 4. MTL prediction of overall $S_{11}(f)$ magnitude response.

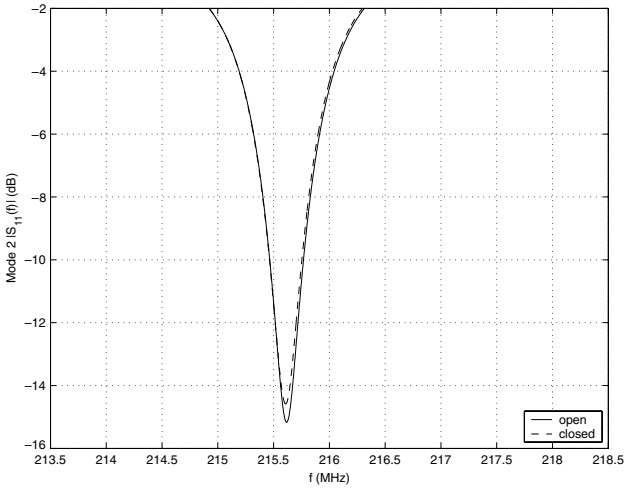

Figure 5. Detailed view of MTL predicted $S_{11}(f)$ magnitude response around second resonant mode.

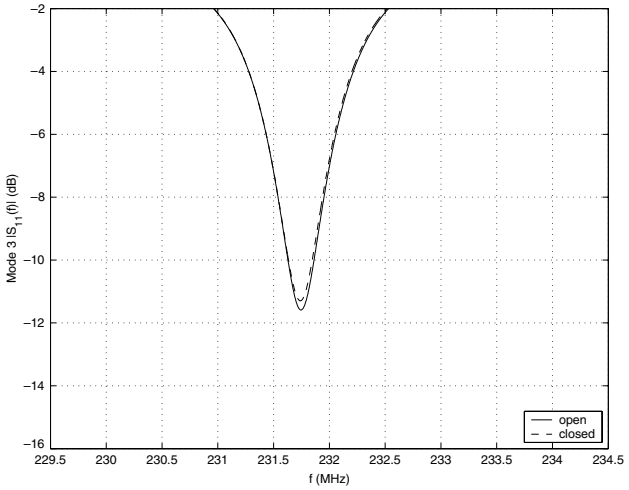

Figure 6. Detailed view of MTL predicted $S_{11}(f)$ magnitude response around third resonant mode.

the MTL model tends to be fairly accurate at predicting the location and depth of the sensor resonances, it is important

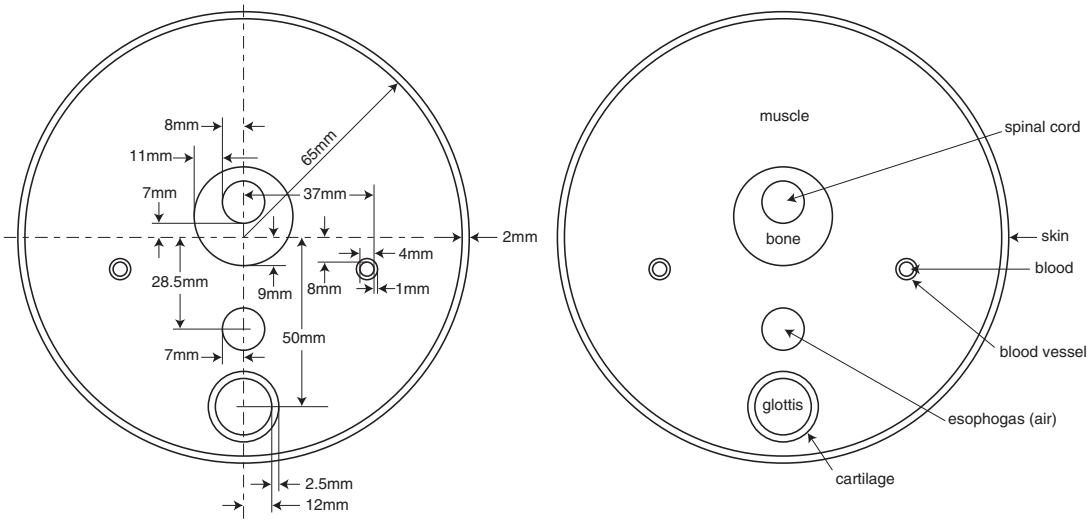

Figure 7. Dimensions and materials of the FE neck model. Note that all structures shown in this figure are cylindrical with a height (in the z-dimension of figure 3) of 38.1 mm. Dielectric properties of the materials are given in table 2.

to point out that the MTL method tends to overestimate the quality factor of the resonances since it does not account for the complex tissue structures and conductive losses encountered in realistic biological systems. This will be addressed with the FE approach in the following.

The results in figure 4 show an almost imperceptible change in the overall $S_{11}(f)$ magnitude response between open and closed glottis states. Nevertheless, a close-up view of the second and third resonances, as shown in figures 5 and 6, reveals the effect of the glottal state more clearly. In the second resonance, MTL analysis predicts a shift in resonant frequency between open and closed glottis states of 8.35 kHz and a shift in $S_{11}(f)$ magnitude of 0.586 dB. In the third resonance, MTL analysis predicts a resonant frequency shift of 5.75 kHz and a $S_{11}(f)$ magnitude shift of 0.292 dB. These results highlight the fact that the TERC sensor must be driven by the RF source at or near a resonant frequency in order to be sensitive to the glottal state.

While the MTL results serve as a useful quick approximation to the behaviour of the sensor, FE analysis can be used with a more elaborate model of the biological load that includes conductive losses in order to confirm the qualitative behaviour of the MTL results and to refine the accuracy of the quantitative predictions. The FE approach, unlike MTL, also accounts for radiation and electromagnetic edge effects.

Figure 7 shows the dimensions and materials used in the FE neck model based on a proportional analysis of Rohen and Yokochi (1993, p 179). The model of the larynx is constructed to include the most significant structures: the skin, the neck muscle, the cartilage around the trachea, the jugular veins, the spinal cord and the vertebrae. The open and closed glottal states are modelled by placing a cylinder of muscle in the 'glottis' region (glottis closed) or a cylinder of air in this region (glottis open). The dielectric properties of the materials of the neck model are given in table 2.

To perform the FE analysis, the load model shown in figure 7 and the TERC sensor described in table 1 were constructed in HFSS[2] version 8.5. Figure 8 shows the overall

[2] HFSS is a registered trademark of Ansoft Corporation.

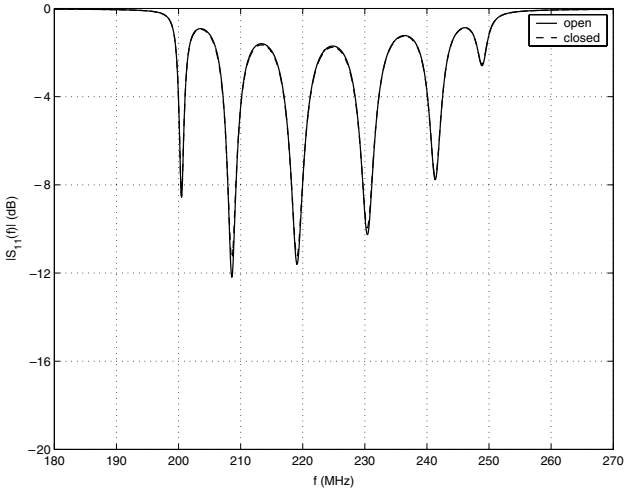

Figure 8. FE prediction of overall $S_{11}(f)$ magnitude response.

predicted $S_{11}(f)$ magnitude response in the open and closed glottis states and figures 9 and 10 show detailed plots of the second and third resonant modes.

As a direct comparison between figure 4 and figure 8 indicates, the FE model predicts broader resonances, i.e. lower quality factors, and a non-unity reflection coefficient at frequencies between the resonances. This is attributable to the more complex and lossy biological load model and the fact that the FE formulation, unlike the MTL approach, can handle radiation and electromagnetic edge effects. A close-up view of the second and third resonances, as shown in figures 9 and 10, reveals the effect of the glottal state. In the second resonance, the FE simulations predict a shift in resonant frequency between open and closed glottis states of 9.5 kHz and a shift in $S_{11}(f)$ magnitude of 0.801 dB. In the third resonance, MTL analysis predicts a resonant frequency shift of 6.5 kHz and a $S_{11}(f)$ magnitude shift of 0.358 dB.

The FE simulations also allow detailed field plots to be generated that show the amount of field penetration into the glottis and provide some intuition as to the sensitivity of the

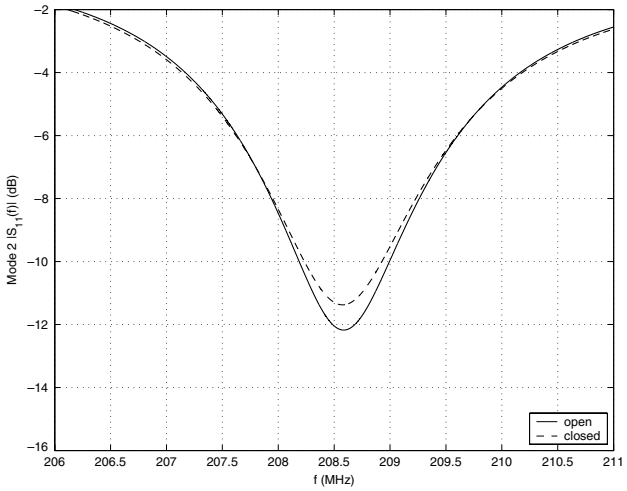

Figure 9. Detailed view of FE predicted $S_{11}(f)$ magnitude response around second resonant mode.

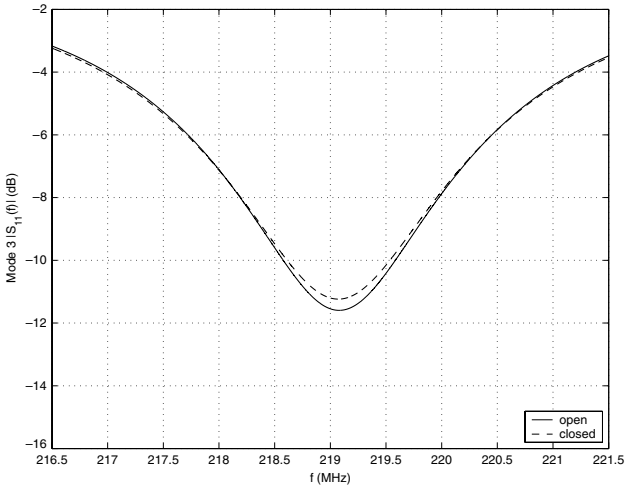

Figure 10. Detailed view of FE predicted $S_{11}(f)$ magnitude response around third resonant mode.

TERC sensor in different modes. Electric field magnitudes are shown for all six resonant modes of the TERC sensor in figures 11–13 for the case when the load is in the 'open glottis' state. These figures confirm the results seen in figures 8–10 where the most sensitive resonant modes have greater electric field magnitudes in the glottis region of the model. Specifically, the field plots indicate that mode 2 has the highest field strength in the glottis which is confirmed by the relatively large change in $S_{11}(f)$ magnitude response observed in mode 2 shown in figure 9. While mode 3 also shows good field penetration, it is not as good as mode 2. This is reflected in the somewhat lower sensitivity to the glottal state for mode 3 shown in figure 10.

4. Experimental results

To confirm the analytical predictions of section 3, a physical prototype of the TERC speech sensor was constructed based on the parameters of table 1. This section describes the construction details of the prototype sensor and the experiments performed with this sensor. Two sets of experimental results are also presented that demonstrate the ability of the TERC speech sensor to measure small vibrations in an acoustically excited phantom load and to measure glottal activity in a human subject in a controlled laboratory environment.

4.1. Prototype construction

A physical prototype of the TERC sensor was constructed primarily of a cylindrical LEXAN former cut to a height of 38.1 mm and with microstrip lines and outer shield formed by placing adhesive-backed copper foil on the inner and outer surfaces of the former. To maximize the quality factor of the sensor, the 108 pF terminating capacitors (consistent with the FE design in section 3.3) were realized by placing four surface-mount parallel-connected BC1308-series capacitors, each with a quality factor of greater than 940, resulting in a total $Q \geqslant 3760$ for the parallel configuration. These capacitor arrays were soldered to the copper foil between the microstrip lines and the outer shield on both the top and bottom edges of the cylindrical former. A standard SMA connector was mounted to the LEXAN former directly adjacent to the end of the shield to allow for ease of connection to a coaxial cable and network analyser. The shell of the SMA connector was soldered directly to the collar shield and the signal pin of the SMA connector was connected to the endmost microstrip though a lumped matching network (as shown in figure 3) to allow for a limited amount of tuning to match to the 50 Ω characteristic impedance of the coaxial cable. Insulating Kapton[3] film was placed over the interior conductors of the TERC sensor and matching circuit to prevent unintended conduction to the load.

In all the experiments described in this section, the TERC speech sensor was connected to port 1 of a Hewlett Packard 8714ES network analyser. The source power was set to 0 dBm and the network analyser was configured to display the logarithmic magnitude of the $S_{11}(f)$ measurement.

4.2. Phantom load experiments

The experiments described in this section all use a non-biological phantom load that closely resembles the dielectric properties of neck tissues over the frequency range of interest. The phantom load is constructed as a cylinder of agarose gel with radius of approximately 65 mm and height of approximately 38 mm. The phantom load was placed inside the TERC speech sensor and the entire assembly was placed on a nonferrous pedestal (a two foot tall LEXAN cylinder with an identical diameter to the TERC sensor prototype) to minimize field distortions due to other objects in the laboratory.

Two experiments were performed with the non-biological phantom load. The first experiment considers the overall $S_{11}(f)$ magnitude response of the sensor in a loaded state with no mechanical vibration to verify that the sensor $S_{11}(f)$ magnitude response agrees with the analytical predictions. Figure 14 shows a plot of the $S_{11}(f)$ magnitude response of the agarose-loaded TERC speech sensor for the frequency range 180 MHz $\leqslant f \leqslant$ 270 MHz. While the depth of the resonances is less than predicted by MTL and FE analysis, the overall qualitative nature of the $S_{11}(f)$ magnitude response

[3] Kapton is a registered trademark of DuPont.

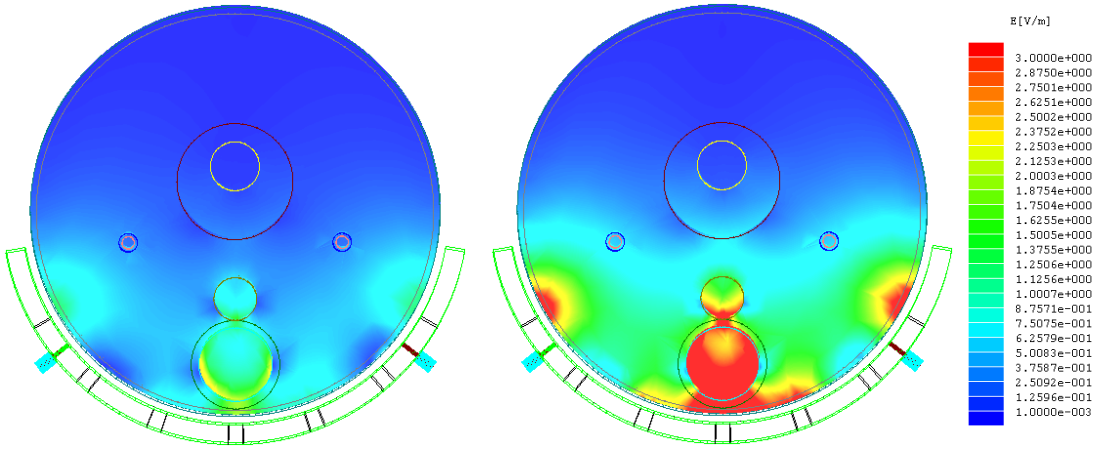

Figure 11. Electric field magnitude in resonant modes 1 and 2.

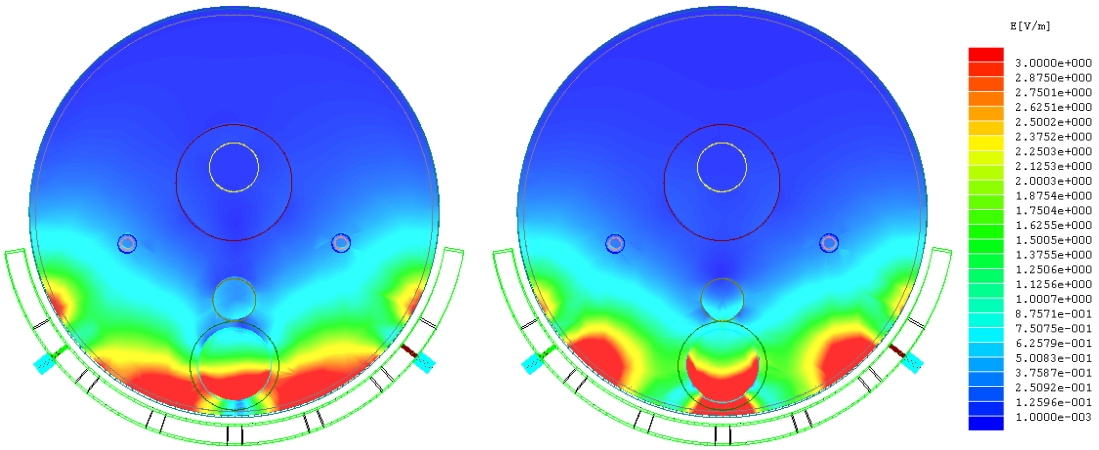

Figure 12. Electric field magnitude in resonant modes 3 and 4.

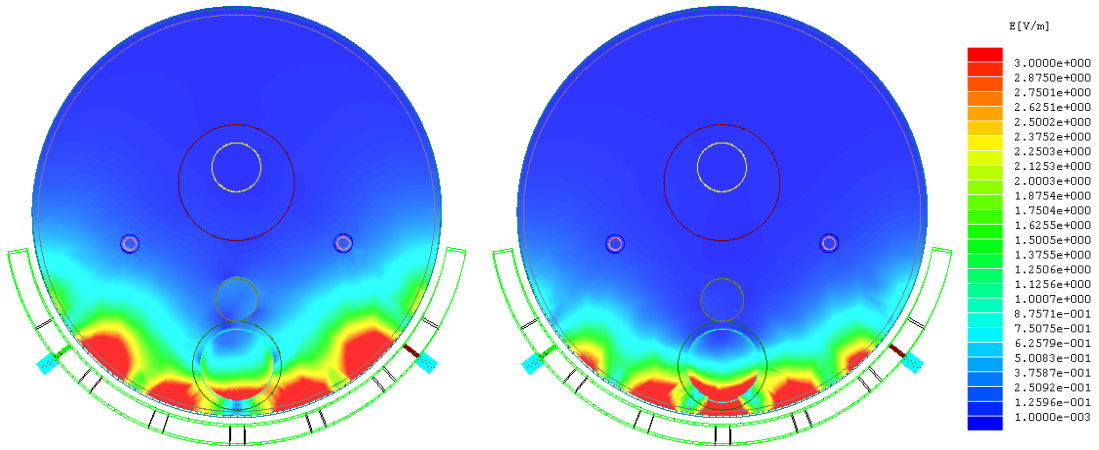

Figure 13. Electric field magnitude in resonant modes 5 and 6.

agrees with the MTL and FE predictions. The experimental $S_{11}(f)$ magnitude response clearly shows the expected six characteristic resonances with the first resonance appearing at approximately 193 MHz.

The second experiment considers the response of the prototype sensor to mechanical vibrations in the non-biological phantom load. We developed a test fixture to acoustically excite the phantom load while placed inside the TERC speech sensor. The agarose gel phantom load was acoustically excited at frequencies in the voiced speech range by a 4 inch diameter loudspeaker placed approximately 2 feet below the gel cylinder. The loudspeaker was driven by an audio power amplifier with a variable frequency sinusoidal source in order to induce small mechanical vibrations though air coupling to the gel cylinder at the excitation frequency. Figure 15 shows the test setup on a laboratory bench.

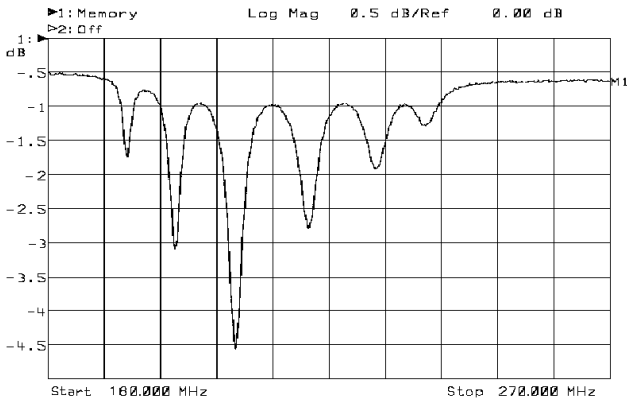

Figure 14. Network analyser S_{11} plot versus frequency of TERC prototype with phantom load (agarose gel cylinder with radius of approximately 68 mm and height of approximately 38 mm).

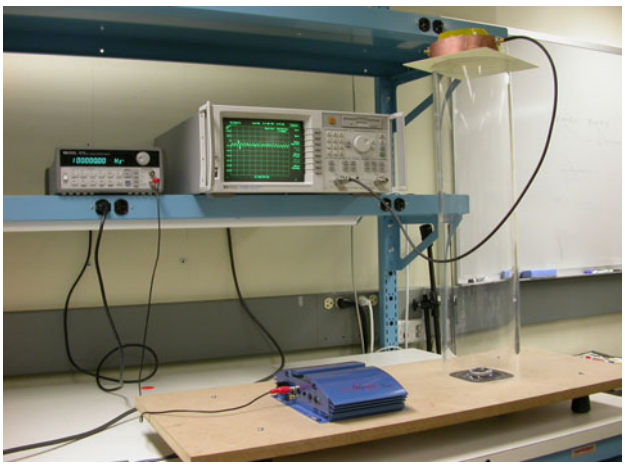

Figure 15. Photograph of the acoustic excitation test fixture used to obtain the results in figure 16.

(This figure is in colour only in the electronic version)

We note that the majority of the coupling between the loudspeaker and the agarose gel is through the air but that a small amount of mechanical coupling may occur through the test fixture as well.

Figure 16 shows a plot of the reflection coefficient of the agarose-loaded TERC sensor versus time where the TERC speech sensor is driven in CW (continuous wave or single frequency) mode. In this experiment, the TERC speech sensor is driven in its third resonant mode (the deepest mode in figure 14). This figure clearly shows a sinusoidal pattern in the reflection coefficient measurements at the 50 Hz and 100 Hz excitation frequencies. To confirm that the sensor was in fact measuring mechanical vibrations in the phantom load and not picking up vibrations in the test fixture or stray fields from the loudspeaker or amplifier, this experiment was repeated with the phantom load removed. The CW S_{11} plot in this case was flat and no sinusoidal excitation was observed.

4.3. Human subject experiments

While the experiments in the prior section demonstrate that the prototype TERC sensor is capable of measuring small mechanical vibrations in a non-biological phantom load, even when weakly matched to the load, human subject experiments are necessary to verify that the TERC sensor is able to detect voiced speech. This section describes a simple human subject experiment using the prototype TERC sensor with the Hewlett Packard 8714ES network analyser.

To allow for sensor placement on a human subject, it was necessary to split the sensor prototype into front and rear semicircular pieces using a band saw. The front piece contained all the conductors, terminating capacitors, the SMA connector and the matching circuit. The rear piece (composed entirely of LEXAN) was not used.

The human subject experiments were conducted in a controlled laboratory environment. While sitting in close proximity to the network analyser, the human subject held the front piece of the prototype sensor in a comfortable position against their neck slightly above the thyroid cartilage (with the aforementioned Kapton film preventing conduction to the subject's skin). The sensor was connected to the network analyser with a coaxial cable. The subject was required to avoid any unnecessary motion for the tests since even slight movements often resulted in relatively large changes to the

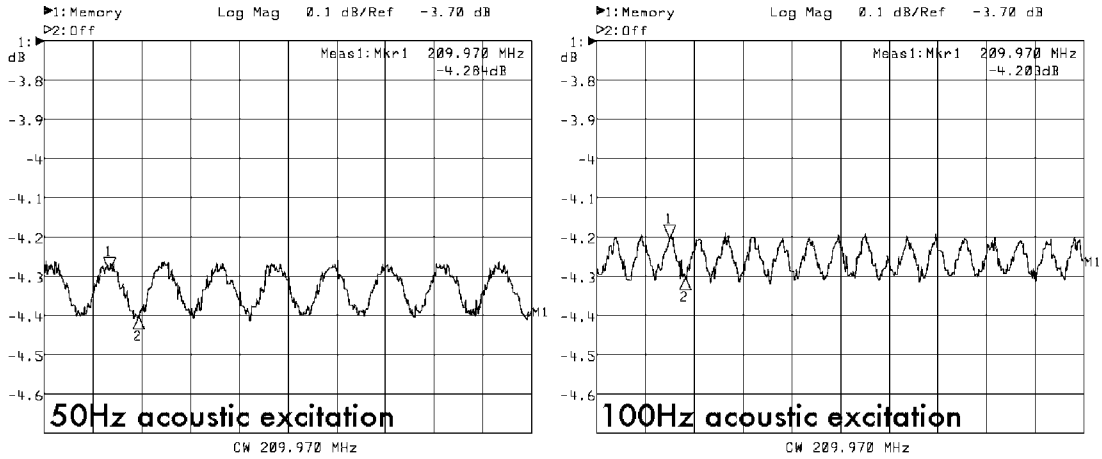

Figure 16. Network analyser S_{11} plots versus time for TERC prototype with a human subject driven in the CW mode at 209.970 MHz (mode 3). The phantom load was acoustically excited at frequencies 50 Hz and 100 Hz. The timespan of each plot is 175 ms.

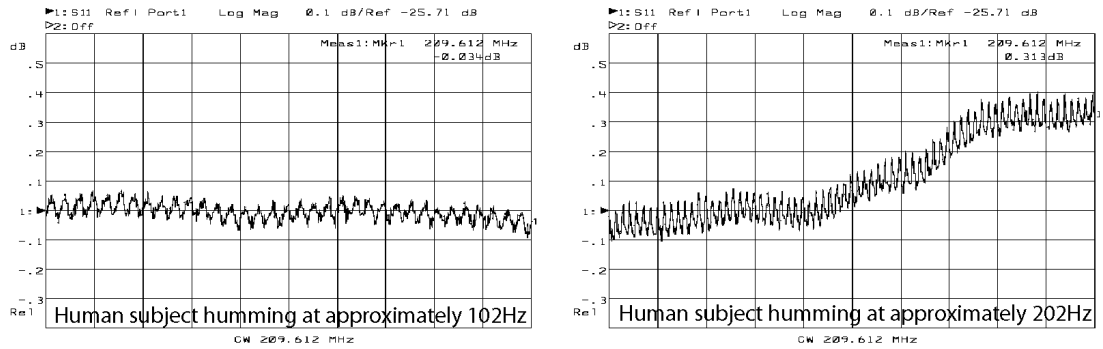

Figure 17. Network analyser S_{11} plots versus time for TERC prototype with a human subject driven in the CW mode at 209.612 MHz (mode 3). The human subject was humming at constant frequencies of approximately 102 Hz and 202 Hz, respectively. The timespan of each plot is 348 ms.

measured $S_{11}(f)$ magnitude response. A laboratory technician tuned the matching network to maximize the depth of the third resonance and then set the network analyser to operate in the CW mode in the third resonance. To induce glottal activity, the subject was instructed to hum a constant tone (to the best of their ability) while a laboratory technician captured the results. A microphone was also used to record the acoustic signal in order to determine the acoustic frequency content and the fundamental frequency of the hummed tone.

Figure 17 shows a plot of the reflection coefficient of the TERC sensor versus time where the TERC speech sensor is driven in the CW mode in the third resonant mode. This figure clearly shows that the $S_{11}(f)$ magnitude response oscillates at a frequency consistent with the fundamental frequency of the hummed tone. While there is some noise present in both results, the fundamental frequency of the measured signal is quite clear and could be used for estimating the fundamental frequency of the hummed tone if it were observed in an acoustically noisy environment. The 202 Hz experimental result also clearly shows a 0.35 dB shift in the $S_{11}(f)$ magnitude response that can be attributed to unintentional subject movement. While this shift is approximately three times larger than actual glottal signal, it can easily be removed with a more sophisticated measurement system since subject movement tends to occur at frequencies well below those of voiced speech.

Finally, to confirm that the sensor was in fact measuring glottal activity, we observed the $S_{11}(f)$ magnitude response while the subject remained silent. The CW S_{11} plot in this case was flat and no oscillation was observed.

5. Conclusions and future research directions

In this paper, we propose a new sensor for non-acoustic voiced speech measurement called the tuned electromagnetic resonator collar. When used in conjunction with a standard network analyser, the TERC sensor is designed to detect small perturbations in the dielectric properties of neck tissue that result from the glottal cycle. A specific sensor design is proposed that leverages recent developments in high-field magnetic resonance imaging. This design allows the use of MTL analytical methods for rapid performance evaluation and optimization. The results of the MTL analysis are presented along with FE simulations and experimental results that confirm the analysis and demonstrate the speech sensing potential of the TERC sensor.

Our primary contribution in this paper is a demonstration of the feasibility of the TERC sensor through analysis, simulation and experimental results. While our experimental results demonstrate that the sensor is able to detect small vibrations in a non-biological load as well as glottal activity in a human subject, a comprehensive human studies test programme is needed to better characterize the speech sensing performance of the TERC sensor. Such a test programme will also show the extent to which the TERC sensor is able to detect speech characteristics beyond glottal activity (such as vocalization) that lend to the overall intelligibility of the speech signal.

There are several practical improvements to the proposed TERC sensor that would help to facilitate an effective human studies test programme. The current TERC sensor design uses a rigid LEXAN former which tends to be uncomfortable for most subjects. One possible solution to this problem would be to use a thin, flexible former to allow the sensor to conform to the talker's neck and to be worn at a desired level of tightness. In addition to improving the comfort of the subject, a flexible former may also help to alleviate the effect of subject movement on the $S_{11}(f)$ magnitude response seen in the current design.

Additional interface circuitry will also need to be developed prior to a comprehensive study of the sensor. The network analyser used for the experiments in section 4 will need to be replaced with an RF signal generator and a demodulation system (similar to the EGG) to allow for longer speech recordings. The demodulation and recording system can also be designed to attenuate baseband frequencies below those of voiced speech to minimize the effects of subject movement.

In the current design, the TERC sensor usually requires some degree of tuning (adjustment of both the matching network and the RF source frequency) for each subject to maximize the speech sensitivity. A typical procedure is to manually tune the matching network while the sensor is worn by the human subject to maximize the depth of the second or third resonance. The RF source is then manually tuned to the strongest resonance of the sensor. Subject motion occasionally necessitates retuning the RF source but rarely necessitates retuning of the matching network. While manual

RF source tuning is possible, albeit inconvenient, in the laboratory environment, automatic RF source tuning will be necessary for field use of the sensor. One possible solution to this problem is the development of an automatic RF resonance tracking system using a phased-locked-loop (PLL) (Gokcek 2003). The bandwidth of the PLL should be designed to be below the minimum frequency of voiced speech. Such a system could, at least in theory, track changes in the sensor's resonant frequency that result from subject movement and automatically maximize the sensitivity of the TERC sensor irrespective of the subject's position.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions as well as Sergey Makarov for his assistance in designing and building matching networks. This work was supported by DARPA/NAVSEA contact number N00024-02-C-6341. Distribution statement A: Approved for public release, distribution unlimited.

References

Baken R 1992 Electroglottography *J. Voice* **6** 98–110

Berenger J 1994 A perfectly matched layer for the absorption of electromagnetic waves *J. Comput. Phys.* **114** 185–200

Bogdanov G and Ludwig R 2002 Coupled microstrip line transverse electromagnetic resonator model for high-field magnetic resonance imaging *Magn. Reson. Med.* **47** 579–563

Boves L and Cranen B 1982 Evaluation of glottal inverse filtering by means of physiological registrations *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* vol 7 pp 1988–91

Brady K, Quatieri T, Campbell J, Campbell W, Brandstein M and Weinstein C 2004 Multisensor melpe using parameter substitution *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) (Montreal, Quebec)* vol 1

Bronzino J 1995 *The Biomedical Engineering Handbook* (Boca Raton, FL: CRC Press)

Burnett G 1999 The physiological basis of glottal electromagnetic micropower sensors (GEMS) and their use in defining an excitation function for the human vocal tract *PhD Thesis* University of California, Davis

Burnett G, Holzrichter J, Gable T and Ng L 1999a Direct and indirect measures of speech articulator motions using low power EM sensors *Proc 14th Int. Congress of Phonetic Sciences (San Francisco, CA)* pp 2247–9

Burnett G, Holzrichter J, Gable T and Ng L 1999b The use of glottal electromagnetic micropower sensors in determining a voiced excitation function *Proc. 138th Meeting of the Acoustical Society of America (Columbus, OH)*

Campbell W, Quatieri T, Campbell J and Weinstein C 2003 Multimodal speaker authentication using nonacoustic sensors *Workshop on Multimodal User Authentication (Santa Barbara, CA)*

Gokcek C 2003 Tracking the resonance frequency of a series *RLC* circuit using a phase locked loop *Proc. 2003 IEEE Conf. on Control Applications (CCA) (Istanbul, Turkey)* vol 1 pp 609–13

Hess W 1983 *Pitch Determination of Speech Signals (Springer Series in Information Sciences)* (New York: Springer)

Holzrichter J, Burnett G, Ng L and Lea W 1998 Speech articulator measurements using low power EM-wave sensors *J. Acoust. Soc. Am.* **103** 622–5

Ludwig R, Bogdanov G, King J, Allard A and Ferris C 2004 A dual RF resonator system for high-field functional imaging of small animals *J. Neurosci. Methods* **132** 125–35

Ng L, Burnett G, Holzrichter J and Gable T 2000 Denoising of human speech using combined acoustic and EM sensor signal processing *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) (Istanbul, Turkey)* vol 1 pp 229–32

Paul C 1994 *Analysis of Multiconductor Transmission Lines* (New York: Wiley Interscience)

Rohen J and Yokochi C 1993 *Color Atlas of Anatomy: A Photographic Study of the Human Body* 3rd edn (New York: Igaku-Shoin Medical Publishers)

Sacks Z, Kingsland D, Lee R and Lee J-F 1995 A perfectly matched anisotropic absorber for use as an absorbing boundary condition *IEEE Trans. Antennas Propag.* **43** 1460–3

Scanlon M 1998 Acoustic sensor for health status monitoring *Proc. IRIS Acoustic and Seismic Sensing* vol 2 pp 205–22

Stevens K 1977 Physics of laryngeal behaviour and larynx modes *Phonetica* **34** 264–79

Titze I 1980 Comments on the myoelastic–aerodynamic theory of phonation *J Speech Hear. Res.* **23** 495–510