Meas. Sci. Technol. **16** (2005) 2381–2390 doi:10.1088/0957-0233/16/11/033

Measuring glottal activity during voiced speech using a tuned electromagnetic resonating collar sensor

D R Brown III[1], **K Keenaghan**[2] **and S Desimini**[3]

[1] Worcester Polytechnic Institute, ECE Department, Worcester, MA, USA
[2] Naval Undersea Warfare Center, Newport, RI, USA
[3] Nortel Networks, Billerica, MA, USA

E-mail: drb@wpi.edu

Received 21 April 2005, in final form 1 August 2005
Published 20 October 2005
Online at stacks.iop.org/MST/16/2381

Abstract
Non-acoustic speech sensors can be employed to obtain measurements of one or more aspects of the speech production process, such as glottal activity, even in the presence of background noise. These sensors have a long history of clinical applications and have also recently been applied to the problem of denoising speech signals recorded in acoustically noisy environments (Ng *et al* 2000 *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP) (Istanbul, Turkey)* vol 1, pp 229–32). Recently, researchers developed a new non-acoustic speech sensor based primarily on a tuned electromagnetic resonator collar (TERC) (Brown *et al* 2004 *Meas. Sci. Technol.* **15** 1291). The TERC sensor measures glottal activity by sensing small changes in the dielectric properties of the glottis that result from voiced speech. This paper builds on the seminal work in Brown *et al* (2004). The primary contributions of this paper are (i) a description of a new single-mode TERC sensor design addressing the comfort and complexity issues of the original sensor, (ii) a complete description of new external interface systems used to obtain long-duration recordings from the TERC sensor and (iii) more extensive experimental results and analysis for the single-mode TERC sensor including spectrograms of speech containing both voiced and unvoiced speech segments in quiet and acoustically noisy environments. The experimental results demonstrate that the single-mode TERC sensor is able to detect glottal activity up to the fourth harmonic and is also insensitive to acoustic background noise.

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

1. Introduction

A promising approach to the problem of denoising speech signals corrupted by strong background noise is the use of non-acoustic speech sensors. A non-acoustic speech sensor measures certain aspects of the speech generation process, e.g. glottal activity, as a proxy for the actual acoustic speech signal and tends to be highly immune to acoustic noise. These sensors have been employed in a variety of applications and have recently been used in conjunction with a microphone and additional signal processing in order to augment the acoustic speech signal and improve speech quality (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 frequently used non-acoustic speech sensor is the accelerometer (Stevens *et al* 1975). An accelerometer is worn in contact with the body of the talker and measures skin or bone vibrations resulting from glottal activity and nasalization. Accelerometers tend to be fairly insensitive to positioning (as long as firm skin contact is maintained) and have been used in a variety of applications including estimating the fundamental frequency of voiced speech (Hess 1983) and estimating the sound pressure level (SPL) of speech (Svec *et al* 2005). Because of their direct mechanical coupling to the talker, accelerometers tend to be insensitive to moderate levels of background noise.

Another well-known non-acoustic speech sensor is the electroglottograph (EGG) (Boves and Cranen 1982, Baken and Orlikoff 2000). Unlike the accelerometer which relies on mechanical coupling, the EGG utilizes low-voltage electrodes on a talker's neck and a sensitive current measurement system to detect changes in electrical impedance across the throat as a proxy for glottal activity during voiced speech.

A more recent example of a non-acoustic speech sensor is the low-power radar-based glottal electromagnetic sensor (GEMS) (Holzrichter *et al* 2005, Burnett 1999). Also a non-mechanical sensor, the GEMS sensor employs a high frequency, e.g. 2 GHz, transmitter/receiver pair and operates as an interferometer where the movement of a reflecting object, e.g. a tracheal wall, can be inferred from phase variations measured in the reflected signal.

While all of these non-acoustic speech sensors have proven to be effective in a laboratory environment, they are subject to certain limitations. The mechanical nature of accelerometers makes them less effective in very high-noise environments where vibrations can be mechanically coupled to the accelerometer through the body of the talker. The EGG tends to be sensitive to the position of the electrodes on the neck where mispositioning the electrodes can lead to a complete loss of signal (Burnett 1999). The radar-based GEMS sensor is also somewhat sensitive to position in the sense that complicated reflective environments may obscure the desired target and lead to ambiguity as to what is actually being measured (Burnett *et al* 1999). The necessity for direct skin contact, inherent to accelerometers as well as the EGG and GEMS sensors, may also prohibit the use of these sensors in some applications.

Recently, a new non-acoustic speech sensor based on magnetic resonance imaging (MRI) technology (Bogdanov and Ludwig 2002, Ludwig *et al* 2004) was developed with the intent of alleviating some of the shortcomings of the existing sensors. The new sensor, first described in Brown *et al* (2004), utilizes a tuned electromagnetic resonator collar (TERC) to measure glottal activity during voiced speech. The TERC sensor is non-mechanical like the EGG and GEMS sensors but, unlike these sensors, does not require direct skin contact due to its capacitive sensing modality. Additionally, with the development of the automatic tuning systems described in Brown *et al* (2004), the TERC sensor's robust design and insensitivity to precise positioning suggests its appropriateness for potential use outside a laboratory environment.

While the analytical, simulation-based and experimental results in Brown *et al* (2004) demonstrated the feasibility of the TERC sensor concept, there were three shortcomings in the prior work that motivate the contributions of this paper. The first shortcoming was that the TERC sensor prototype was complicated to construct and was also uncomfortable for a subject to wear due to its rigid construction. The discomfort of wearing the TERC sensor made experimental work difficult and precluded the use of the sensor outside a controlled laboratory setting. The second shortcoming was that a network analyser was used to obtain all of the experimental results reported in Brown *et al* (2004). The network analyser limited the scope of the experiments to short-duration humming tests and limited the scope of the analysis to simple time-domain plots. Finally, the third shortcoming was that no experiments with actual speech, with or without background noise, were reported; only the results of two humming tests in a quiet environment were presented. As such, the sensing and noise rejection capabilities of the sensor could not be characterized beyond the preliminary results presented in Brown *et al* (2004).

This paper addresses each of the shortcomings in the original work. The primary contributions of this paper are (i) a description of an improved single-mode TERC sensor design addressing the comfort and complexity issues of the original sensor, (ii) a complete description of new external interface systems used to obtain long-duration recordings from the TERC sensor and (iii) more extensive experimental results and analysis for the single-mode TERC sensor including spectrograms of speech containing both voiced and unvoiced speech segments in quiet and acoustically noisy environments. The experimental results suggest that the single-mode TERC sensor is able to measure glottal activity up to the fourth harmonic and also effectively rejects strong acoustic background noise.

2. TERC sensor concept and principle of operation

Like the EGG and GEMS sensors, the TERC sensor is a device used for measuring the glottal activity that occurs during voiced speech. The source of acoustic energy for all voiced speech is the glottal cycle (Stevens 1977, 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, and 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 speech signal produced at the output of the vocal tract. This property of speech production is exploited by almost all non-acoustic voiced speech sensors including accelerometers which measure skin vibrations primarily induced by glottal activity.

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 of the order of 50–200 times that of air for frequencies in the range 10–200 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

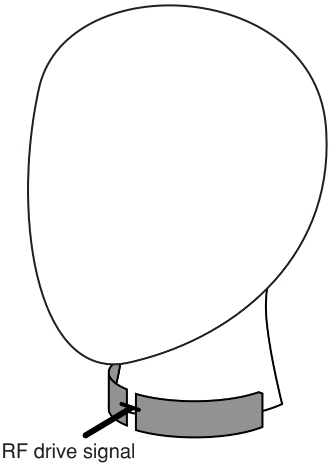

Figure 1. TERC sensor concept.

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 speech signal.

One method for measuring the relative permittivity of a particular material is to build a capacitor with known dimensions and the desired material in its interior. As an example, consider a simple parallel plate capacitor with a time-varying relative 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. In this case, since the capacitance is proportional to permittivity, measuring changes in the capacitance provides a straightforward method for measuring changes in the permittivity.

In order to measure glottal activity through this capacitive modality, a capacitor must be constructed with the glottal region of the talker in its interior. Figure 1 illustrates the TERC sensor approach to this problem. Although the TERC sensor is not a parallel plate capacitor, the previously developed intuition still applies: changes in the relative permittivity in the interior of the sensor are manifested as changes in the sensor's capacitance.

Because the glottis represents only a small fraction of the aggregate dielectric in the interior of the TERC sensor, the changes in capacitance that occur during voiced speech are typically quite small (Brown *et al* 2004) and the accuracy of typical capacitance measurement methods may not be sufficient to clearly distinguish the glottal cycle. Hence, a fundamental challenge with the approach shown in figure 1 is the development of a method by which the small changes in capacitance resulting from glottal activity can accurately be measured. One approach to this problem is to measure the reflection coefficient of the sensor, defined as

$$S_{11}(f) = \frac{Z_{\text{in}}(f) - Z_0}{Z_{\text{in}}(f) + Z_0}, \tag{2}$$

where $Z_{\text{in}}(f)$ is the complex input impedance of the sensor at frequency f and Z_0 is the characteristic impedance of the coaxial cable providing the RF drive signal. The TERC sensor's complex input impedance is a function of the sensor's design parameters, the frequency f, and the effective relative permittivity of the material in the sensor's interior (which is itself a function of the frequency f). While the glottal cycle typically causes only small perturbations to $Z_{\text{in}}(f)$, these perturbations can result in large relative changes to $S_{11}(f)$ if $|S_{11}(f)| \approx 0$ or, equivalently, if $|Z_{\text{in}}(f) - Z_0| \approx 0$. In other words, if the sensor can be designed with a deep resonance at some frequency f, small changes in the permittivity of the material in the interior of the sensor can be observed 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 a capacitive sensor to always maintain a deep resonance at a consistent frequency 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 capacitive plates of the TERC sensor and the RF source. These lumped elements are tuned while the sensor is worn by the talker to optimize the depth and quality factor of the sensor's resonance.

Once the tuning of the matching network is complete, the TERC sensor can be employed as a glottal activity sensor. The TERC sensor is driven by an RF source at low power (typically

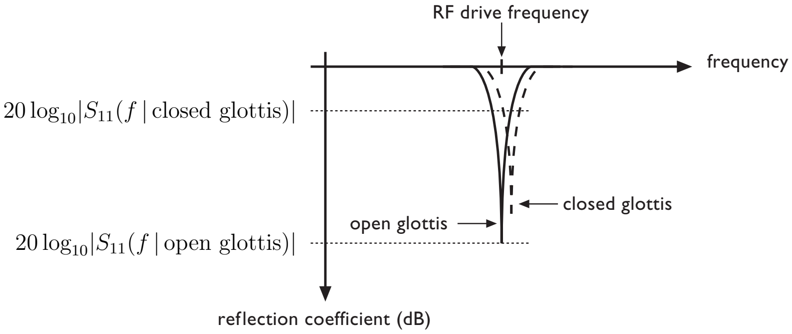
RF drive frequency
frequency
$20\log_{10}|S_{11}(f\,|\,\text{closed glottis})|$
closed glottis
open glottis
$20\log_{10}|S_{11}(f\,|\,\text{open glottis})|$
reflection coefficient (dB)

Figure 2. Illustration of the sensitivity enhancement principle of the TERC sensor where small changes in the permittivity of the material in the interior of the sensor result in relatively large changes in the sensor's reflection coefficient.

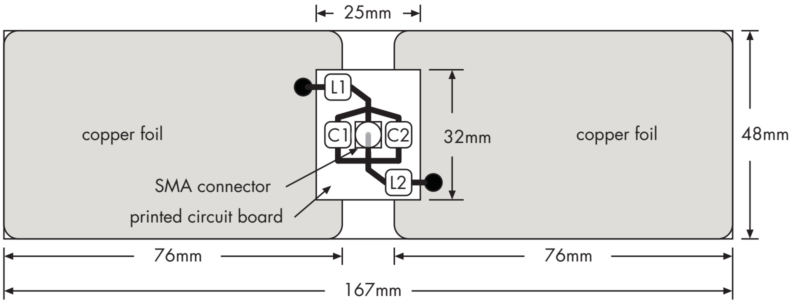

Figure 3. Dimensions of the single-mode TERC sensor prototype and layout of the balanced passive matching network. All capacitors and inductors shown are tunable.

−10 dBm or less) at a frequency close to the sensor's resonant frequency. Voiced speech causes small changes in the TERC sensor's complex input impedance which are manifested as relatively large changes in the amplitude of the reflected signal from the TERC sensor. The envelope of the reflected signal represents the talker's glottal waveform and is recovered by employing an amplitude demodulator, as described in section 4.

3. Single-mode TERC sensor prototype construction

Based on preliminary simulations using finite element analysis (see Brown *et al* (2004) for a description of the tools and methods), a prototype of the improved single-mode TERC sensor was constructed as shown in figure 3. The single-mode TERC sensor's design is considerably less complicated than the original TERC sensor developed in Brown *et al* (2004). The original TERC sensor was constructed on a rigid former with seven foil strips and possessed a multitude of resonances. The finite element analysis in Brown *et al* (2004) showed that the first resonant mode was the most sensitive mode to glottal activity and, consequently, the higher resonances of the original TERC sensor were not used. The simplified design of the single-mode TERC sensor shown in figure 3 uses a flexible former with only two foil strips resulting in only one resonant mode.

The single-mode TERC sensor[4] was constructed on a 0.5 mm thick substrate of flexible clear acrylic plastic. Two appropriately sized strips were cut from a sheet of adhesive-backed 0.1 mm thick copper foil and affixed to the acrylic substrate. The printed circuit board containing an SMA connector as well as a balanced passive matching network was then centred on the substrate and the appropriate connections were soldered to the copper foil strips. The printed circuit board used in the prototype was 1.57 mm thick and had traces printed on both the top and bottom layers. The SMA connector serves as the single connection from the sensor to the drive and demodulation circuitry described in section 4. The balanced matching network of tunable capacitors and inductors used to tune the resonance of the TERC sensor for each subject consisted of two Coilcraft model 142-09J08 inductors with a tuning range of 0.47–0.57 μH and two Sprague–Goodman GCL40000 tunable plastic dielectric capacitors with a tuning range of 1–40 pF.

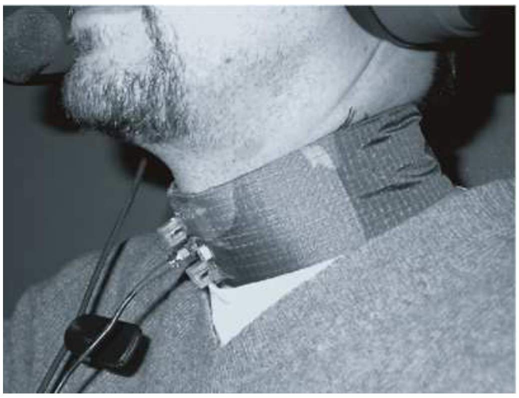

Figure 4. Photograph of the placement of the single-mode TERC sensor.

Two steps were also taken to increase the durability and comfort of the TERC sensor without compromising its function. Initial sensor prototypes occasionally experienced solder fractures at the connection between the rigid printed circuit board and the flexible foil strips. This problem was resolved by securing the printed circuit board to the sensor with a bead of electronic grade silicone rubber adhesive sealant. By applying the sealant to the entire perimeter of the printed circuit board, the durability of the prototype was significantly improved and no subsequent solder fractures were observed. The comfort of the sensor was improved by enclosing the sensor in a nylon fabric sleeve with a Velcro closure. Holes were cut in the sleeve to allow for the coaxial cable connection to the SMA connector and also for tuning of the matching network. A photograph of the sensor in its sleeve as worn by a male subject is shown in figure 4.

4. TERC sensor drive and demodulation circuitry

In Brown *et al* (2004), experimental results were obtained by connecting the TERC sensor to a network analyser which generated a low-power continuous wave RF drive signal and directly measured the time-varying reflection coefficient of the sensor. While the network analyser provided a convenient solution to the drive and demodulation systems necessary to the operation of the TERC sensor, its signal capture capabilities were limited to short-duration tests and its analysis capabilities were limited to time-domain plots. This section describes new TERC sensor drive and demodulation systems developed to

[4] For brevity, the 'single-mode TERC sensor' will be referred to as the 'TERC sensor' hereafter.

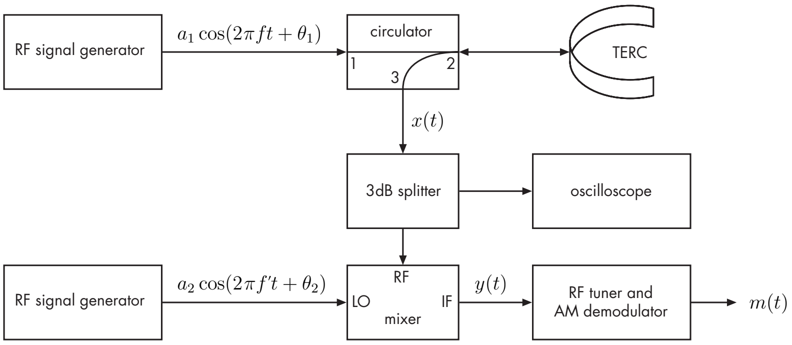

Figure 5. Block diagram of the TERC sensor drive and demodulation systems.

facilitate extended speech recordings and more sophisticated signal analysis.

Figure 5 shows an overview of the TERC sensor drive and demodulation circuitry developed to obtain the experimental results described in section 5. After the TERC sensor's matching network has been tuned[5], and prior to the commencement of speech testing, the RF signal generator must be set to a frequency close to the TERC sensor's resonant frequency. The frequency of the drive signal is held constant over each test but may be adjusted slightly between experiments due to small changes in subject's position and the consequent changes in the TERC sensor's resonance. The 3 dB splitter and oscilloscope at the output of the circulator are used to facilitate these adjustments. The test technician finds the resonant frequency of the TERC sensor by adjusting the drive frequency until the amplitude of the reflected signal is minimized.

The circulator shown in figure 5 was designed to provide unity transmission and high isolation over the expected range of TERC sensor resonant frequencies (Wenzel 2003). Transmission was measured to be within ±0.2 dB and isolation was measured to be in excess of 25 dB between 20 MHz and 80 MHz. All ports of the circulator were designed with a characteristic impedance of approximately 50 Ω to avoid signal reflections.

As discussed in section 2, the envelope of the reflected signal from the TERC sensor represents the glottal waveform of the talker. The circulator captures this reflected signal at port 2 and transmits it to port 3. The signal at port 3 of the circulator can be written as

$$x(t) = a_3[1 + \alpha m(t)] \cos(2\pi f t + \theta_3), \tag{3}$$

where a_3 is the amplitude of the reflected signal, $m(t)$ is the unit-amplitude baseband glottal waveform, α is the effective amplitude modulation (AM) index, f is the drive frequency and θ_3 is the phase offset. Because the TERC sensor's modulation index α is typically less than 0.02 (Keenaghan 2004), a sensitive AM demodulator is required to recover the baseband glottal waveform $m(t)$. The PC-based WinRadio[6] WR-G303i software-defined receiver was tested and determined to provide satisfactory performance in these operating conditions. The tuning range of the WR-G303i receiver is, however, limited to 9 kHz–30 MHz which is below the typical resonant frequency range for the TERC sensor shown in figure 3. To overcome this limitation, a second RF signal generator, set to frequency f', and a passive RF mixer were used to generate a mixed signal prior to demodulation. The mixed signal can be written as

$$y(t) = a_4[1 + \alpha m(t)]\{\cos(2\pi(f - f')t + \theta_4) \\ + \cos(2\pi(f + f')t + \theta_4)\} \tag{4}$$

where the frequency of the mixing signal is set so that $f - f' \approx 20$ MHz. The high-frequency component of the mixed signal in (4) is automatically rejected by the WR-G303i's built-in coherent RF tuner and no low-pass filter is necessary prior to demodulation. The baseband output of the WR-G303i receiver representing the glottal waveform $m(t)$ was sampled and recorded, as described in the following section, for analysis and post-processing.

5. Experimental methods and results

This section describes experimental results obtained with a male subject wearing a prototype of the TERC sensor as described in section 3 with drive and demodulation circuitry as described in section 4. The experimental methodology is first outlined and then followed by results demonstrating the ability of the TERC sensor to measure glottal activity while rejecting acoustic background noise.

5.1. Experimental methodology

A male subject was recruited for all of the experiments described in this section. The subject was seated in a low-resonance listening booth with equipment arranged as shown in figure 6. Inside the listening booth, a reference microphone was placed on a stand with a shock mount approximately 25 cm in front of the subject. The subject was instrumented with both the TERC sensor (placed on the subject as shown in figure 4) and a physiological microphone[7] affixed to the centre of the subject's forehead with a Velcro strap. The output signals from all three sensors were routed outside the listening booth and the radio-frequency TERC sensor signal was demodulated

[5] As discussed in section 5, a network analyser was used to manually tune the matching network of the TERC sensor for each subject prior to speech testing. Once the matching network was satisfactorily tuned, the TERC sensor was connected to the drive and demodulation systems shown in figure 5.

[6] http://www.winradio.com.

[7] The physiological microphone is a piezoelectric accelerometer with a gel capsule for efficient skin coupling. The first prototype of this device was described in Scanlon (1998) and is now commercially available from BIOPAC Systems, Inc. at http://www.biopac.com.

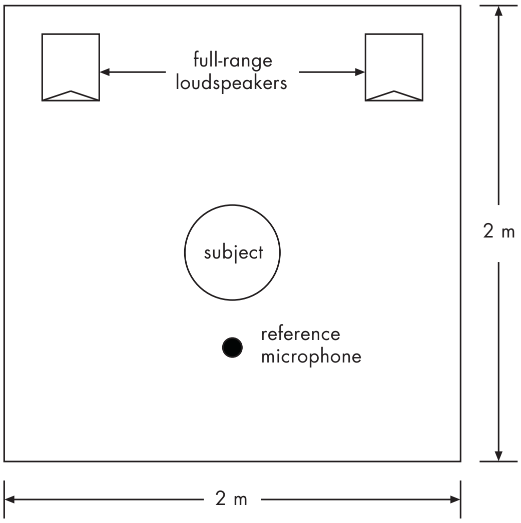

Figure 6. Arrangement of the low-resonance listening booth for all speech tests. The subject is seated facing the reference microphone. The loudspeakers and microphone are positioned at approximately the same height as the seated subject's head.

as described in section 4 to produce a baseband audio signal. The audio signals of all three sensors were then separately amplified to avoid any potential crosstalk and synchronously sampled with a multichannel recording interface at 44.1 kHz with 16 bits of resolution. The digital output of the recording interface was connected to a personal computer and the signals were recorded into separate uncompressed audio files with a professional multichannel audio-recording software package for analysis and post-processing.

Prior to testing, a network analyser was used to tune the matching network of the TERC sensor such that the depth of the sensor's resonance was maximized. Figure 7 shows an example of a typical $S_{11}(f)$ sweep after tuning. Resonance depths in excess of −30 dB were typically achieved with the TERC sensor and matching network described in section 3. Once tuned, the subject was instructed to avoid unnecessary motion and the RF drive signal was adjusted to match the TERC sensor's resonant frequency.

A total of three speech tests were performed. The first test required the subject to speak a list of ten Harvard psychoacoustic sentences (Keenaghan 2004). This test was performed in order to evaluate the capabilities of the TERC sensor over a variety of voiced and unvoiced speech segments. The Harvard psychoacoustic sentence test required approximately 1 min to complete and was performed while the subject was in a quiet environment (22 dBA SPL). The second and third tests required the subject to speak a list of 14 individual vowel-emphasized words in quiet and high-noise (98 dBA SPL) environments, respectively. The vowel-emphasized word list was selected in order to facilitate a comparison between the quiet and high-noise environments over long uninterrupted segments of voiced speech. The second and third tests each took approximately 40 s to complete. The subject was retuned, if necessary, between tests and was required to repeat a test if any of the sentences or words were incorrectly spoken.

In the third test, the high-noise environment was achieved by playing recordings of military vehicle interior cabin noise at high volume levels through the loudspeakers placed behind the subject in the listening booth. The power spectral density (PSD) of the military vehicle background noise (as recorded by the reference microphone in the listening booth with the subject present) is shown in figure 8. To avoid potential hearing damage (and to facilitate communications between the subject and the test engineer stationed outside the listening booth), the subject wore a commercial aviation headset with passive noise attenuation of approximately 24 dB.

5.2. *Harvard psychoacoustic sentence test results*

Unlike the humming tests described in Brown *et al* (2004), Harvard psychoacoustic sentences require the subject to generate both voiced and unvoiced speech and are a test of the TERC sensor's ability to detect various types of speech. Figure 9 shows the spectrograms of the recorded signals[8] for all three sensors for the first sentence in the list: 'A yacht slid around the point into the bay'. The results shown in figure 9 were typical of the other nine sentences in the test and, consistent with the results described in Brown *et al* (2004), demonstrate the ability of the TERC sensor to detect glottal activity during voiced speech. While the bandwidth of the TERC sensor measurements is not as broad as the reference microphone or the physiological microphone, the detailed spectrogram shown in figure 10 demonstrates that the TERC sensor is able to clearly distinguish the fundamental frequency as well as several harmonics of the glottal waveform during voiced speech. As expected, the TERC sensor does not appear to measure unvoiced speech segments.

Three recording artefacts were also present in the results shown in figure 10. A 60 Hz noise component, consistent with the mains frequency in the United States where these tests were performed, is clearly visible in the physiological microphone recording. This noise is presumed to be due to imperfect shielding of the physiological microphone circuitry and/or one or more cables. The second and third artefacts appear in the TERC sensor recordings. These recordings contain a strong low-frequency noise component which typically peaks at approximately 8 Hz. This noise is attributed to unavoidable subject movement as well as the behaviour of the TERC sensor's AM demodulation system at very low modulation indices. A simple high-pass filter could be employed to suppress this inaudible low-frequency noise without adversely affecting the quality of the TERC sensor's measurements. In addition to the low-frequency noise, a 75 Hz tone is also faintly but consistently visible in all of the TERC sensor recordings. This tone appears to be an anomaly of the WR-G303i AM demodulation system due to its presence irrespective of the test, the background acoustic noise environment (see figures 11 and 12), and whether the subject was or was not speaking. This measured power in this tone was typically 12 dB or more below the power in the fundamental glottal frequency.

[8] Note that all of the signals presented in this section and in section 5.3 are original recordings and have not been filtered or otherwise modified other than gain normalization to facilitate comparisons.

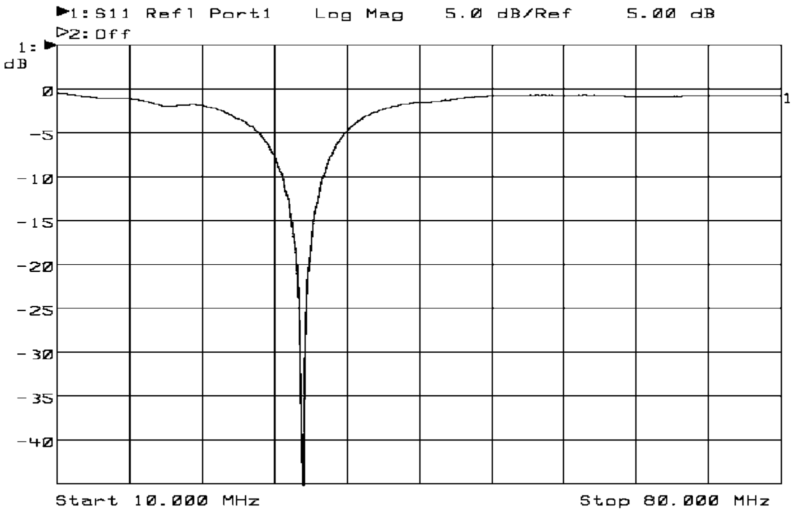

Figure 7. Typical TERC sensor $S_{11}(f)$ sweep after tuning of the matching network.

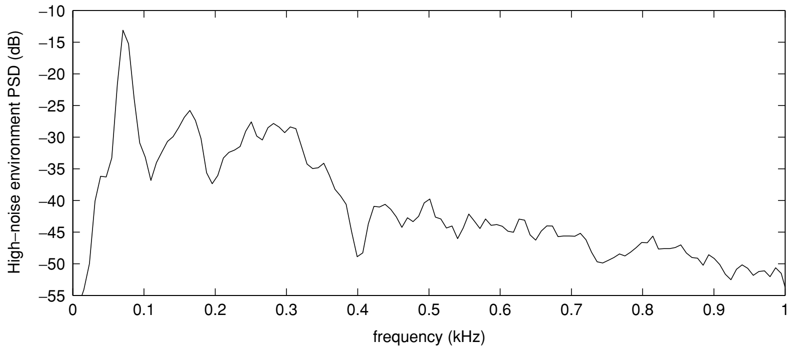

Figure 8. The power spectral density of the military vehicle noise recorded by the reference microphone in a high-noise environment experiment.

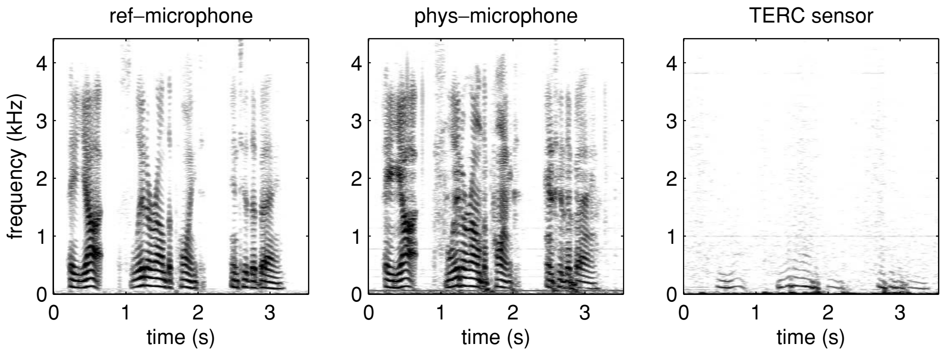

Figure 9. Spectrograms of the sentence ‘A yacht slid around the point into the bay’ for the reference microphone, physiological microphone and TERC sensor in the quiet environment. Note the 175 ms (approximate) delay inherent in the TERC measurements due to the digital demodulation circuitry.

5.3. Vowel-emphasized word test results

After completing the Harvard psychoacoustic sentence tests, the subject was instructed to read and speak a list of 14 preselected single-syllable words, pausing between each word and extending the vowel portion of each word. This test was performed in the quiet environment and then repeated in the high-noise environment. Figure 11 shows detailed spectrograms of the recorded signals for all three sensors in both noise environments. Two consecutively spoken vowel-emphasized words, ‘pear’ and ‘join’, are shown. The effect of the strong background acoustic noise can be clearly seen in the reference microphone and physiological microphone recordings in the high-noise environment where the speech is almost completely obscured below 400 Hz. The TERC sensor, on the other hand, appears to be unaffected by the background acoustic noise. Glottal activity up to the fourth harmonic is clearly visible in both TERC sensor recordings

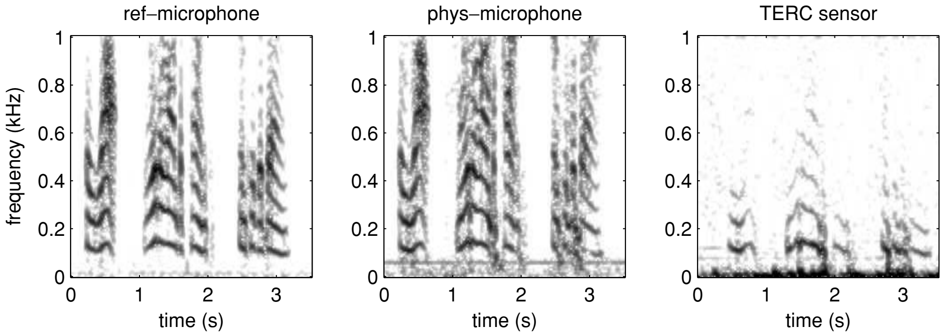

Figure 10. Detailed view of the spectrograms shown in figure 9 showing the harmonic structure of the TERC sensor's measurements.

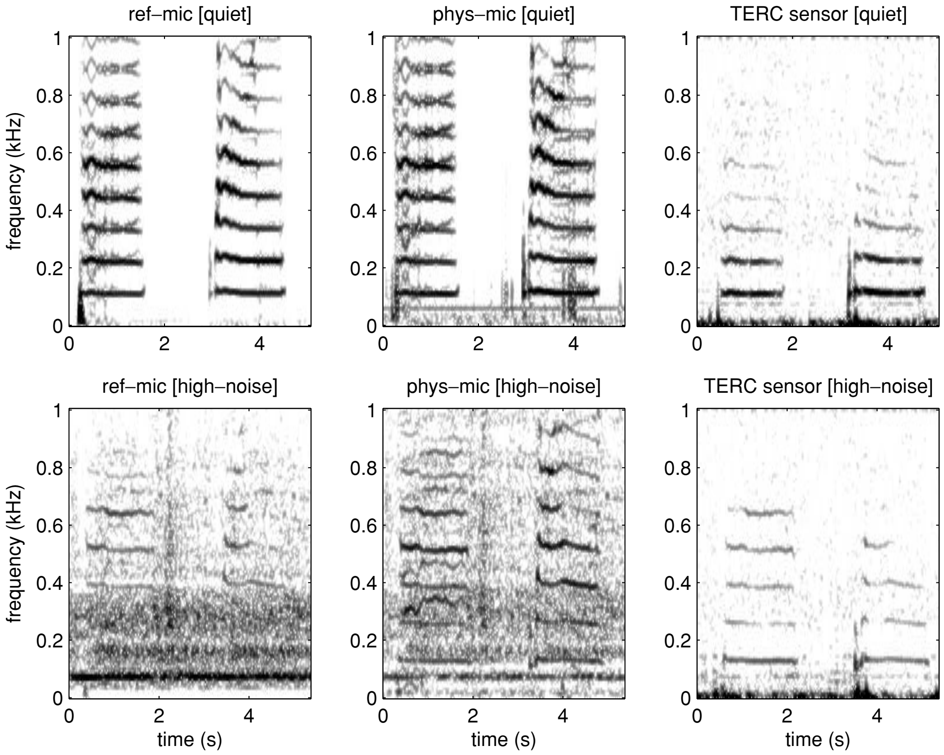

Figure 11. Detailed spectrograms of the vowel-emphasized words 'pear' and 'join' for the reference microphone, physiological microphone and TERC sensor in both the quiet and high-noise environments.

and the background noise appears to be completely rejected. The only indicator of the strong background noise present in the high-noise environment TERC sensor recording is the raised pitch and different harmonic structure of the speech with respect to the quiet environment. This can be attributed to the Lombard reflex (Junqua 1993) as the subject was inclined to raise both their speech volume and pitch in the high-noise environment despite the use of passive hearing protection in the experiments.

To directly compare the performance of the reference microphone, physiological microphone and TERC sensor, figure 12 plots the power spectral density of a 1 s snapshot of the vowel-emphasized portion of the word 'pear' for all three sensors in both noise environments. These results show that all three sensors perform well in the quiet environment but that only the TERC sensor is able to clearly distinguish the fundamental frequency and lower harmonics of the speech signal in the high-noise environment. To quantitatively compare the performance of the sensors over the entire high-noise recording, we also computed the magnitude of the PSD of each sensor's output at the fundamental speech frequency (typically between 110 Hz and 150 Hz) and the magnitude of the PSD of each sensor's output at the peak noise frequency (approximately 73 Hz) for each of the 14 vowel-emphasized words. Denoting the ratio of these two quantities as the 'fundamental-to-peak-noise ratio' (FPNR), table 1 shows the results of these calculations. The mean FPNR results are consistent with the results shown for the specific case in figure 12. Moreover, the *worst-case* FPNR of the TERC sensor over all 14 words was 23 dB better than the *best-case* FPNR of the reference microphone and 5 dB better than the *best-case* FPNR of the physiological microphone. These results quantitatively demonstrate, at least for this particular data set, that the TERC

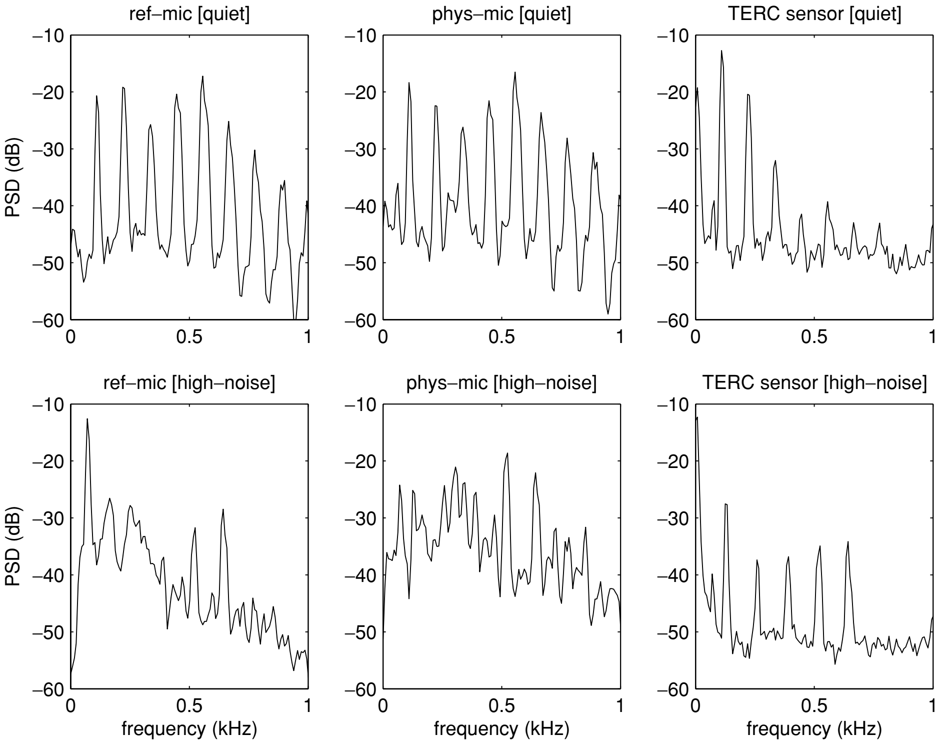

Figure 12. Comparison of the reference microphone, physiological microphone and TERC sensor speech recordings in the quiet and high-noise environments for the vowel-emphasized portion of the word ‘pear’.

Table 1. Fundamental to peak noise ratio calculations for the 14 words in the vowel-extended noise test in the high-noise environment.

	Minimum FPNR (dB)	Maximum FPNR (dB)	Mean FPNR (dB)
Reference microphone	−32	−13	−16.8
Physiological microphone	−7	5	1.0
TERC sensor	10	20	14.7

sensor is able to consistently extract accurate fundamental frequency information of voiced speech corrupted by high levels of background noise.

6. Conclusions and future research directions

This paper builds upon the original work in Brown *et al* (2004) and addresses several shortcomings in the original sensor design, the supporting systems and the scope of the experimental results. In this paper, we present (i) a description of an improved single-mode TERC sensor design addressing the comfort and complexity issues of the original sensor, (ii) a complete description of new external interface systems used to obtain long-duration recordings from the TERC sensor and (iii) more extensive experimental results and analysis for the single-mode TERC sensor including spectrograms of speech containing both voiced and unvoiced speech segments in quiet and acoustically noisy environments. The experimental results suggest that the single-mode TERC sensor is able to measure glottal activity up to the fourth harmonic and also effectively rejects strong acoustic background noise.

While the experimental results presented in this paper demonstrate the capabilities and potential of the TERC sensor, there are two aspects of the current TERC sensor design that require additional development before this technology can be effectively employed outside a controlled laboratory environment. The TERC sensor itself is highly portable, robust and inexpensive to construct, but the supporting drive and demodulation systems shown in figure 5 are bulky, somewhat fragile and expensive. The components in the drive and demodulation systems are, however, mostly general-purpose instruments that provide significantly more functionality than necessary. Special-purpose hardware will need to be developed to improve the size, robustness and cost of the TERC sensor’s drive and demodulation systems.

The TERC sensor, unlike the physiological microphone, also currently exhibits high sensitivity to subject movement. This sensitivity is, however, somewhat different from the positional sensitivity exhibited by the EGG. The TERC sensor has been used to obtain clear glottal signals from voiced speech tests with the sensor worn in a variety of positions on the neck. In fact, no special effort was made to control the position of the sensor in the experiments presented in this paper other than to rotate the sensor so that it was approximately symmetric around the trachea. Our experiments suggest that, unlike the EGG, the TERC sensor is relatively insensitive to the position in which it is worn once the matching network and drive frequency have been appropriately tuned.

The TERC sensor’s sensitivity to subject movement appears to be due primarily to the fact that the sensor’s matching network and drive frequency are manually tuned to optimize the sensitivity for the subject's current position. These settings remain fixed for the duration of each experiment and, because of the sharpness of the sensor's resonance, even small unintentional subject movements during the experiment can lead to a loss of signal from the TERC sensor and require subsequent retuning. While manual tuning is mostly an inconvenience in a controlled laboratory setting, it is a more significant problem in less controlled environments with unrestricted talker movement. To reduce the TERC sensor's sensitivity to subject movement, an automated resonance tracking system, such as the one described in Gokcek (2003), will need to be developed. This system is designed to track changes in the sensor's resonant frequency that result from subject movement and automatically adjust the RF drive signal to continuously maximize the sensitivity of the TERC sensor.

Acknowledgments

This work was supported by DARPA/NAVSEA contact number N00024-02-C-6341.

References

Baken R and Orlikoff R 2000 *Clinical Measurement of Speech and Voice* 2nd edn (San Diego, CA: Singular)

Bogdanov G and Ludwig R 2002 Coupled microstrip line transverse electromagnetic resonator model for high-field magnetic resonance imaging *Magn. Reson. Med.* **47** 579–63

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, Canada)* vol 1

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

Brown D III, Ludwig R, Pelteku A, Bogdanov G and Keenaghan K 2004 A novel non-acoustic voiced speech sensor *Meas. Sci. Technol.* **15** 1291–302

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 1999 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 F, Ng L C, Burke G J, Champagne N J, Kallman J S, Sharpe R M, Kobler J B, Hillman R E and Rosowski J J 2005 Measurements of glottal structure dynamics *J. Acoust. Soc. Am.* **117** 1373–85

Junqua J 1993 The lombard reflex and its role on human listeners and automatic speech recognizers *J. Acoust. Soc. Am.* **93** 510–24

Keenaghan K 2004 A novel non-acoustic voiced speech sensor: experimental results and characterizaton *Master's Thesis* Worcester Polytechnic Institute. Available online at http://www.wpi.edu/Pubs/ETD/Available/etd-0114104-144946/

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

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 behavior and larynx modes *Phonetica* **34** 264–79

Stevens K, Kalikow D and Willemain T 1975 A miniature accelerometer for detecting glottal waveforms and nasalization *J. Speech Hear. Res.* **18** 594–9

Svec J, Titze I and Popolo P 2005 Estimation of sound pressure levels of voiced speech from skin vibration of the neck *J. Acoust. Soc. Am.* **117** 1386–94

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

Wenzel C 2003 Low frequency circulator/isolator uses no ferrite or magnet. Available online at http://www.wenzel.com/pdffiles/RFDesign3.pdf