Aliph™

Theory and use of the Aliph RadioVibrometer (ARV)

Version 1.3

Gregory C. Burnett, Ph.D.

1485 Bayshore Blvd.
Suite 382, Mail Stop 93
San Francisco, CA 94124
www.aliph.com

Table of contents

1. Introduction

The Aliph RadioVibrometer (ARV), formerly known as the Aliph General Electromagnetic Motion Sensor (Aliph GEMS), is a homodyne radiofrequency interferometer capable of very accurate vibration detection and measurement of a reflective surface with a few centimeters of the ARV antenna. These measurements can be made in air or in human tissue, depending on the antenna and method used to take the measurements. The ARV emits only about a milliwatt of power at 2.4 GHz and is compliant with all FDA and FCC requirements regarding safe use for human tissue measurements.

In this paper, we will cover the theory of the operation of the ARV and its use in human subjects tests associated with speech coding and noise suppression. The ARV is also being studied for use in swallowing treatments and other movement and vibration experiments, but we will restrict ourselves to the use of the ARV to detect vibrations caused by speech on and in the neck, cheek, and nose. We will discuss the ways to record and process the ARV signal, as well as ways to use the ARV signal to calculate voice activity, pitch, voiced excitation functions, and the transfer function of the vocal tract. Most of these quantities are available solely through the use of the ARV, and in addition are immune to environmental acoustic noise. Some applications also require the speech output to be recorded, but even these applications benefit from the acoustic noise immunity of the ARV.

Finally, we will discuss the limitations of the ARV and its associated hardware. The ARV is a research tool and requires careful handling to get reliable data. However, the unique information available from the ARV offers many opportunities for the researcher willing to take the time to understand how it works and what it has to offer.

2. Homodyne phase interferometry vibration detection

A homodyne phase interferometer detects moving objects by transmitting electromagnetic (EM) waves at a target and mixing the reflected waves with the original transmitted wave. This mixing results in an output amplitude that is proportional to the phase difference between the mixed waves. Any interaction with the fields of the antenna (including both the near and far fields) can affect this phase difference and result in an amplitude and phase change of the ARV output signal. For stability, very low frequencies are filtered, but if the motion is above about 20 Hz, the amplitude of the ARV signal will vary directly with the interaction of the target(s) with the fields of the ARV antenna.

2.1 Homodyne phase interferometry

We will first examine how the ARV operates in the far field. Let the carrier frequency of the ARV be denoted by f_c. The emitted wave is then

$$v_e(t) = E_0 \sin(2\pi f_c t + \phi)$$

where E_o is the initial amplitude of the wave and ϕ is a general phase offset. For simplicity, we assume for this section that the target is a flat surface perpendicular to the antenna and that it is located far enough away to only interact with the far field. (Figure 2.1). If r(t) describes the motion of the target, then the phase of the reflected wave will be

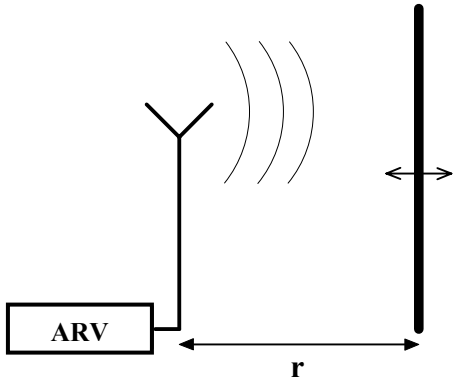

Figure 2.1. Motion of a reflective plate with respect to the ARV.

$$\phi_R(t) = \frac{2\pi}{\lambda}(2r(t)) + \pi = \frac{4\pi}{\lambda} r(t) + \pi .$$

The extra π is due to the 180-degree phase shift that normally occurs when reflecting from a conducting surface or from a smaller-to-greater dielectric constant medium. We are neglecting any change to the frequency due to the

velocity of the target, which we will justify at the end of this section. The reflected wave is described as

$$v_R(t) = RE_0 \sin\left(2\pi f_c t + \phi + \phi_e + \phi_R(t)\right)$$

where R is the reflection coefficient (< 1) and ϕ_e is a constant phase delay due to the sensor hardware. This wave is then mixed with the emitted wave (I) and the emitted wave delayed by 90 degrees (Q). Thus

$$v_I(t) = \alpha_I A \sin\left(2\pi f_c t + \phi + \phi_e + \phi_R(t)\right) \cdot \sin\left(2\pi f_c t + \phi\right)$$
$$v_Q(t) = \alpha_Q A \sin\left(2\pi f_c t + \phi + \phi_e + \phi_R(t)\right) \cdot \cos\left(2\pi f_c t + \phi\right)$$

where $A = RE_0^2$ and α_I and α_Q are the gains of the **I** and **Q** modulation circuits. Ideally, these gains would be equal to each other, but there may be an imbalance, which we also will examine below. Using the identities

$$\sin A \cos B = \frac{1}{2}\left(\sin(A - B) + \sin(A + B)\right)$$
$$\sin A \sin B = \frac{1}{2}\left(\cos(A - B) - \cos(A + B)\right)$$

we get

$$v_I(t) = \frac{A\alpha_I}{2}\left(\cos\left(\phi_e + \phi_R\right) - \cos\left(4\pi f_c t + 2\phi + \phi_e + \phi_R(t)\right)\right)$$
$$v_Q(t) = \frac{A\alpha_Q}{2}\left(\sin\left(\phi_e + \phi_R\right) + \sin\left(4\pi f_c t + 2\phi + \phi_e + \phi_R(t)\right)\right)$$

Using simple LP filters to remove the high frequency second terms, we get

$$v_I(t) = \frac{A\alpha_I}{2}\left(\cos\left(\phi_e + \phi_R(t)\right)\right)$$
$$v_Q(t) = \frac{A\alpha_Q}{2}\left(\sin\left(\phi_e + \phi_R(t)\right)\right)$$

or

$$v_I(t) = \frac{A\alpha_I}{2}\cos\left(\phi_e + \frac{4\pi}{\lambda} r(t) + \pi\right)$$
$$v_Q(t) = \frac{A\alpha_Q}{2}\sin\left(\phi_e + \frac{4\pi}{\lambda} r(t) + \pi\right) \qquad \textbf{(2.1)}$$

To recover the original it may appear that we may simply take the inverse sin and cosine:

$$r(t) = \frac{\lambda}{4\pi}\left(\arccos\left(\frac{2v_I(t)}{A\alpha_I}\right) - (\phi_e + \pi)\right)$$
$$r(t) = \frac{\lambda}{4\pi}\left(\arcsin\left(\frac{2v_Q(t)}{A\alpha_Q}\right) - (\phi_e + \pi)\right)$$

However, this assumes we know ϕ_e, which is unlikely. However, if the position signal r(t) in (1) is oscillatory:

$$r(t) = r_0 + B\sin(2\pi f_r t + \phi_r)$$

(where the small *r* denotes position rather that the large *R* of reflection) then Equation 2.1 becomes

$$\begin{aligned} v_I(t) &= \frac{A\alpha_I}{2}\cos\left((\phi_e + \pi) + \frac{4\pi}{\lambda}(r_0 + B\sin(2\pi f_r t + \phi_r))\right) \\ v_Q(t) &= \frac{A\alpha_Q}{2}\sin\left((\phi_e + \pi) + \frac{4\pi}{\lambda}(r_0 + B\sin(2\pi f_r t + \phi_r))\right) \end{aligned} \qquad \textbf{(2.2)}$$

These equations are quite similar to those used in frequency modulation (for example, see [1] and [2]). The top equation in Equation 2.2 can then be expanded into its fundamental and harmonics using Bessel functions of the first kind:

$$v_I(t) = v_0 + v_1 + v_2 + \ldots$$

where

$$v_0 = \frac{A\alpha_I}{2}\cos\left((\phi_e + \pi) + \frac{4\pi r_0}{\lambda}\right)\cdot J_0\left(\frac{4\pi B}{\lambda}\right)$$

and

$$v_N = A\alpha_I\left[\begin{aligned}&\cos\left((\phi_e + \pi) + \frac{4\pi r_0}{\lambda}\right)\cdot J_{2N}\left(\frac{4\pi B}{\lambda}\right)\cos(2N\cdot 2\pi f_r t) \\ &\qquad - \sin\left((\phi_e + \pi) + \frac{4\pi r_0}{\lambda}\right)J_{2N-1}\left(\frac{4\pi B}{\lambda}\right)\sin((2N-1)\cdot 2\pi f_r t)\end{aligned}\right]$$

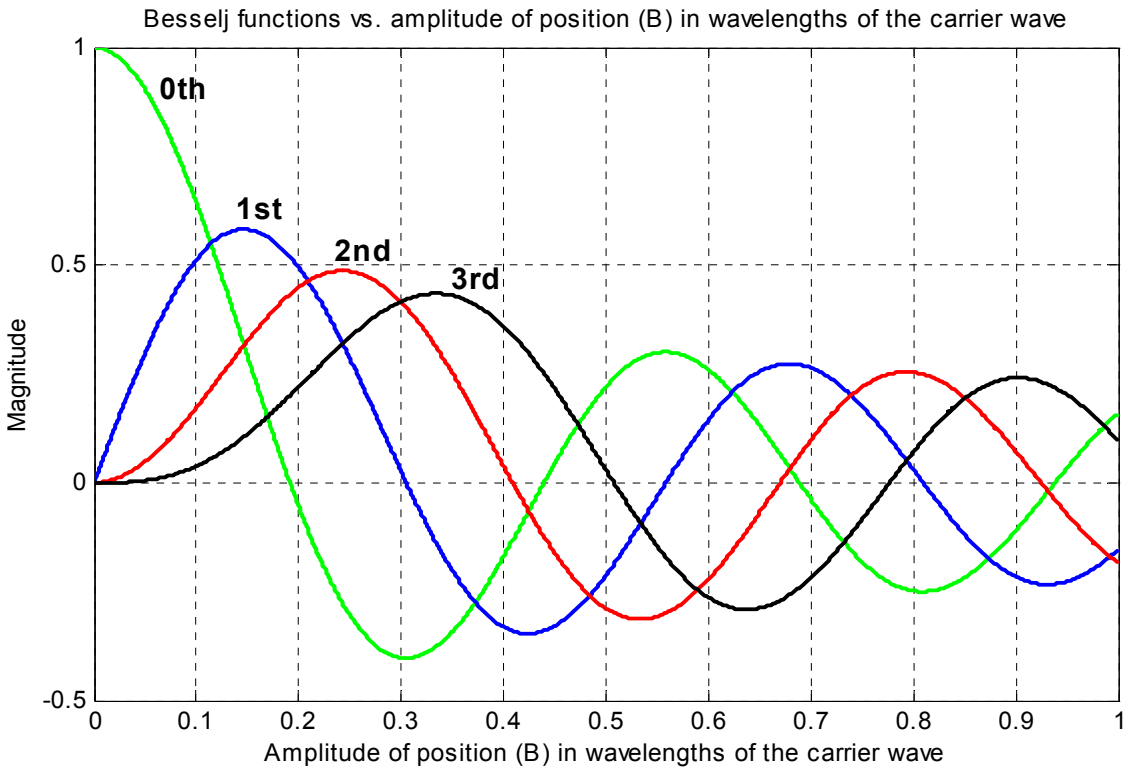

Figure 2.2. Bessel functions of the first kind vs. amplitude of motion of the target in wavelengths of the carrier wave. A value of 1 wavelength means the amplitude of motion of the target is equal to the wavelength of the carrier wave. The amplitude of each Bessel function is the amplitude of the respective harmonic, so the idea is to keep the amplitude of the first order Bessel function much higher than that of the others. The zeroth order function is included for completeness but is not a factor as DC is filtered out of the system.

where $J_N(\beta)$ is the N^{th} Bessel function of the first kind evaluated at β. The first four Bessel functions are plotted in Figure 2.2. For ease of notation, the fundamental will be referred to as the "first harmonic" and DC the "zeroth" harmonic. For N = 1, the $\cos(4\pi f_r t)$ term is the amplitude of the second harmonic (double the fundamental frequency), and $\sin(2\pi f_r t)$ the amplitude of the fundamental. In this way, with increasing N, one can add as many harmonics as needed to make the approximation a good one. Since

$$\cos(x+\pi) = -\cos(x)$$
$$\sin(x+\pi) = -\sin(x)$$

the amplitude for each harmonic for the **I** modulation is

$$DC = \frac{-A\alpha_I}{2}\cos\left(\phi_e + \frac{4\pi r_0}{\lambda}\right)\cdot J_0\left(\frac{4\pi B}{\lambda}\right) \qquad \textbf{(2.3)}$$

$$N^{th}\text{ harmonic} = (-1)^{N+1} A\alpha_I \sin\left(\phi_e + \frac{4\pi r_0}{\lambda} + \frac{M\pi}{2}\right) J_N\left(\frac{4\pi B}{\lambda}\right)\sin\left(2\pi(Nf_r)t + \frac{M\pi}{2}\right)$$

where N = 1, 2, 3… and M = remainder(N-1,2). The harmonics are known as "sidetones" in frequency modulation. The DC component is filtered out, but the harmonics are not and can significantly distort the signal.

Let's examine the first few harmonics for each modulation, assuming:

1. A = 1 (simple units renormalization)
2. α_I and α_Q are approximately unity (simplification, no loss of generality)
3. ϕ_e is a small constant = 0.1 (no loss of generality)
4. $r_o = 5\lambda / 4$ (object is about 15.5 cm away in air, or 2.5 cm in tissue)
5. $B = \lambda / 80$ (amplitude of the target is much smaller than a wavelength, at 2.4 GHz in air this would be an amplitude of about $\pm$ 1.5 mm)
6. $\phi_r = 0$ (no loss of generality)

Then, for the **I** modulation:

$$1^{st}\text{ harmonic} = \sin(0.1+5\pi)J_1\left(\frac{\pi}{20}\right)\sin(2\pi f_r t) = (-0.10)(0.078)\sin(2\pi f_r t) = -0.0078\sin(2\pi f_r t)$$

$$2^{nd}\text{ harmonic} = -\cos(0.1+5\pi)J_2\left(\frac{\pi}{20}\right)\cos(2\pi 2f_r t) = -(-0.995)(0.0031)\cos(2\pi 2f_r t) = 0.0031\cos(2\pi 2f_r t)$$

$$3^{rd}\text{ harmonic} = \sin(0.1+5\pi)J_3\left(\frac{\pi}{20}\right)\sin(2\pi 3f_r t) = (-0.10)(8.06\text{x}10^{-5})\sin(2\pi 3f_r t) = -8.05\text{x}10^{-6}\sin(2\pi 3f_r t)$$

This means that the top equation in 2.2 can now be approximated by (using the above assumptions)

$$v_I(t) \approx -0.0078\sin(2\pi f_r t) + 0.0031\cos(2\pi 2f_r t) - 8.05\text{x}10^{-6}\sin(2\pi 3f_r t) + \ldots$$

Thus we have simplified Equation 2.2 to a series of the position signal (with a different magnitude) added to its harmonics. It is clear that for this expansion, the second harmonic (twice the fundamental) is 270 degrees out of phase and fully half the amplitude of the first harmonic. This addition of a substantial

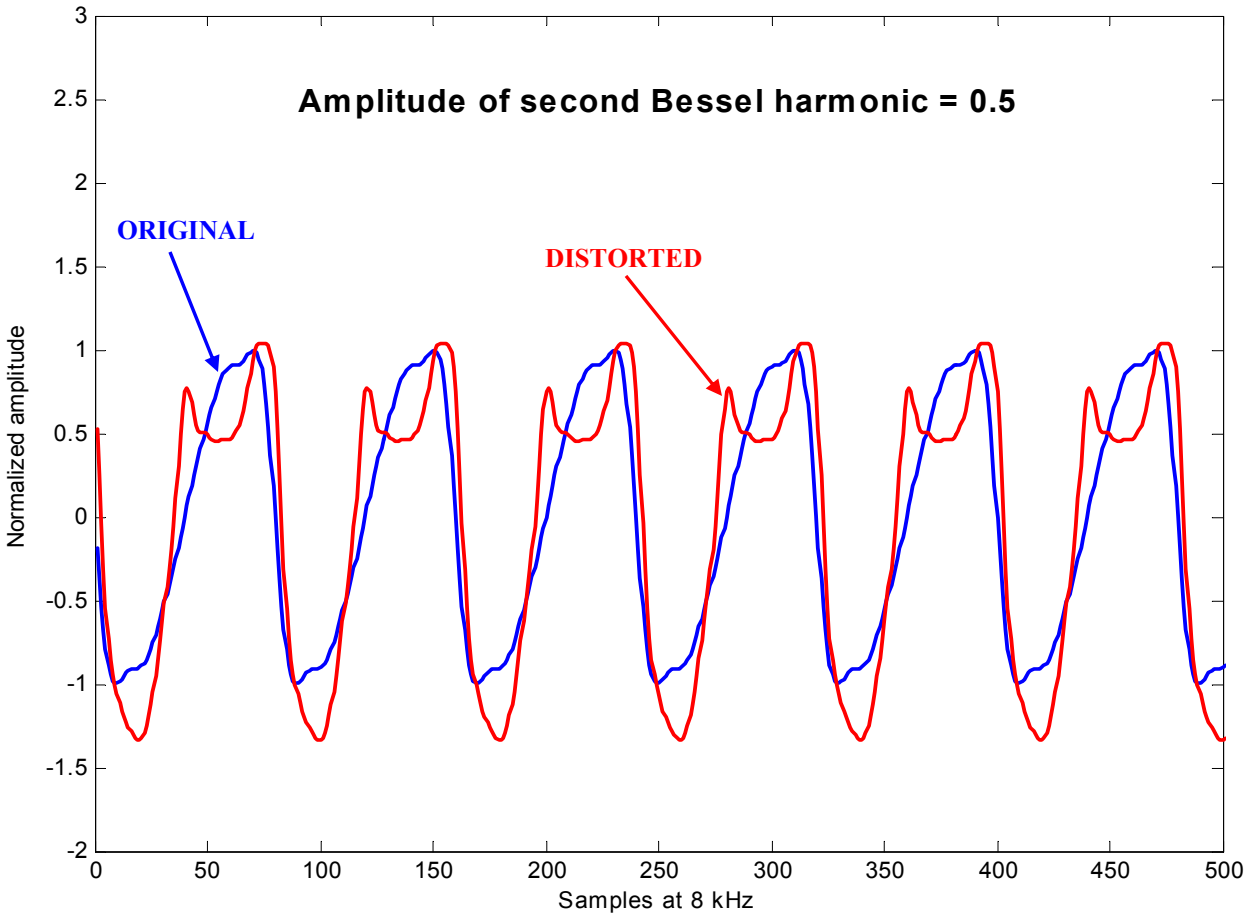

Figure 2.3. Distortion of a complex (fundamental plus four overtones, amplitudes [1, 0.22, 0.14, 0.09, 0.04]) signal when the amplitude of the second Bessel harmonic is 0.5 as in the **I** modulation example above. The original signal is blue and the distorted signal is red. Clearly the distortion of the signal is significant and has led to sharp features with no basis in the original signal. Fortunately, the corresponding **Q** modulation signal would have a second Bessel harmonic of only 0.005, which results in essentially no distortion of the desired position signal.

second harmonic results in the distortion of the position signal (see Figure 2.3). The third harmonic (and all higher orders for this example) is of negligible amplitude and does not contribute significantly to the distortion.

We may follow the same path with the **Q** modulation, with ϕ_e replaced by $\phi_e + \pi/2$. Thus the second half of Equation 2.2 becomes

$$v_Q(t) = \frac{A\alpha_Q}{2}\cos\left(\left(\phi_e + \frac{\pi}{2}\right) + \frac{4\pi}{\lambda}\left(r_0 + B\sin\left(2\pi f_r t + \phi_r\right)\right)\right)$$

and, since

$$\cos\left(x+\frac{\pi}{2}\right)=-\sin(x)$$

$$\sin\left(x+\frac{\pi}{2}\right)=\cos(x)$$

the result is

$$\mathrm{DC}=\frac{-A\alpha_Q}{2}\sin\left(\phi_e+\frac{4\pi r_0}{\lambda}\right)\cdot J_0\left(\frac{4\pi B}{\lambda}\right) \qquad \textbf{(2.4)}$$

$$N^{th}\text{ harmonic}=A\alpha_Q\cos\left(\phi_e+\frac{4\pi r_0}{\lambda}+\frac{M\pi}{2}\right)J_N\left(\frac{4\pi B}{\lambda}\right)\sin\left(2\pi(Nf_r)t+\frac{M\pi}{2}\right)$$

with N = 1, 2, 3… and M = remainder(N-1,2) as before. The harmonics for the Q modulation are therefore

$$1^{st}\text{ harmonic}=\cos(0.1+5\pi)J_1\left(\frac{\pi}{20}\right)\sin(2\pi f_r t)=(-0.995)(0.078)\sin(2\pi f_r t)=-0.078\sin(2\pi f_r t)$$

$$2^{nd}\text{ harmonic}=\sin(0.1+5\pi)J_2\left(\frac{\pi}{20}\right)\cos(2\pi 2f_r t)=(-0.10)(0.0031)\cos(2\pi 2f_r t)=-0.00030\cos(2\pi 2f_r t)$$

$$3^{rd}\text{ harmonic}=\cos(0.1+5\pi)J_3\left(\frac{\pi}{20}\right)\sin(2\pi 3f_r t)=(-0.995)(8.06\text{x}10^{-5})\sin(2\pi f_r t)=-8.02\text{x}10^{-5}\sin(2\pi 3f_r t)$$

resulting in

$$v_Q(t)\approx -0.078\sin(2\pi f_r t)-0.00030\cos(2\pi 2f_r t)-8.02\text{x}10^{-5}\sin(2\pi 3f_r t)+\ldots$$

which is similar to the **I** modulation, except the signs of the harmonics no longer alternate and the first harmonic (fundamental) is ten times as large and the second harmonic (unwanted overtone) is ten times smaller, so the SNR of the **Q** modulation in this example would be 100 times = 40 dB greater than the **I** modulation. This is the reason for using **I/Q** demodulation – when one modulation (the **I** in this case) results in a distorted signal; the other experiences almost no distortion. The phase of the result may be off by 180 degrees

(depending on the distance r_0), but this does not change the frequency characteristics or the location of fast events (i.e. pulses).

*Most importantly, If both the **I** and **Q** modulations are not recorded, the signal will be significantly distorted for some target distances and amplitudes, and there is no way to determine that the signal is distorted.* This can lead to erroneous conclusions based upon distorted signals.

Fortunately, if the gain of each channel is about the same, the channel with the least distortion is also the one with the largest energy. Therefore a simple energy comparison between channels can be made at any time to determine the correct (least distortion) channel to use. We may also use statistical combining methods, as discussed in Section 5.5.

Finally, if there is no 180 degrees phase shift at the reflection interface (due to the dielectric constant of the initial medium being less than the second medium, or perhaps due to near-field interaction) the results for **I** and **Q** above are both simply multiplied by a negative sign – all of the above conclusions are still valid.

The important thing to get from this section is that distortions can be found in either channel, but normally not both at the same time. The severity of the distortions is determined by the round trip distance traveled by the transmitted wave and the relative amplitude of the target vibration.

2.2 Small-amplitude non-sinusoidal motion

It is also possible to describe the ARV signal for non-sinusoidal motion[1]. Starting from Equation 2.1 above:

[1] Thanks to Dr. Samuel L. Earp of Multisensor Science LLC, for assistance in this section.

$$v_I(t) = \frac{A\alpha_I}{2}\cos\left(\phi_e + \frac{4\pi}{\lambda}r(t) + \pi\right)$$
$$v_Q(t) = \frac{A\alpha_Q}{2}\sin\left(\phi_e + \frac{4\pi}{\lambda}r(t) + \pi\right) \quad \textbf{(2.1)}$$

and assuming that

$$r(t) = r_0 + \tilde{r}(t)$$

where r_0 is again a constant term and $\tilde{r}(t)$ is an oscillatory (but not necessarily sinusoidal) motion, we can rewrite the above relationships using trigonometric relationships as

$$v_I(t) = \frac{A\alpha_I}{2}\left[\cos(\psi)\cos\left(\frac{4\pi}{\lambda}\tilde{r}(t)\right) - \sin(\psi)\sin\left(\frac{4\pi}{\lambda}\tilde{r}(t)\right)\right]$$
$$v_Q(t) = \frac{A\alpha_Q}{2}\left[\sin(\psi)\cos\left(\frac{4\pi}{\lambda}\tilde{r}(t)\right) + \cos(\psi)\sin\left(\frac{4\pi}{\lambda}\tilde{r}(t)\right)\right]$$

where

$$\psi = \phi_e + \frac{4\pi}{\lambda}r_0 + \pi.$$

If we assume that the motion $\tilde{r}(t)$ is small with respect to a fraction of a wavelength:

$$\max\left(\left|\tilde{r}(t)\right|\right) << \lambda/4\pi$$

then

$$\cos\left(\frac{4\pi}{\lambda}\tilde{r}(t)\right) \approx 1$$
$$\sin\left(\frac{4\pi}{\lambda}\tilde{r}(t)\right) \approx \left(\frac{4\pi}{\lambda}\tilde{r}(t)\right)$$

and the above results simplify to

$$v_I(t) \approx \frac{A\alpha_I}{2}\left[\cos(\psi) - \sin(\psi)\left(\frac{4\pi}{\lambda}\tilde{r}(t)\right)\right]$$
$$v_Q(t) \approx \frac{A\alpha_Q}{2}\left[\sin(\psi) + \cos(\psi)\left(\frac{4\pi}{\lambda}\tilde{r}(t)\right)\right].$$

The first terms are DC and easily filtered out, leaving us with

$$v_I(t) \approx -A\alpha_I \sin(\psi)\left(\frac{2\pi}{\lambda}\tilde{r}(t)\right)$$
$$v_Q(t) \approx A\alpha_Q \cos(\psi)\left(\frac{2\pi}{\lambda}\tilde{r}(t)\right) \qquad \textbf{(2.5)}$$

This clearly illustrates the quadrature nature of the two channels and shows that as long as the amplitude of motion is small compared to $\lambda/4\pi$ (max amplitude of about 0.01λ), the motion will be faithfully reproduced (except for any effects due to the filtering action of the ARV). This is similar to the results derived from the Bessel expansion above, but does not allow a description of the distortions observed when the amplitude of motion is not small compared to a wavelength of the transmitted wave.

Incidentally, both the Bessel expansion and the small-amplitude approximation above converge if we include only the fundamental frequency in Equation 2.3:

$$v_I(t) = A\alpha_I \sin\left(\phi_e + \frac{4\pi r_0}{\lambda}\right) J_1\left(\frac{4\pi B}{\lambda}\right) \sin(2\pi(f_r)t) \qquad \textbf{(2.3), 1st harmonic}$$

We have included only the **I** modulation for conciseness, the **Q** modulation is analogous. If we assume (as above) that the amplitude of motion B << $\lambda/4\pi$, then the Bessel function reduces to

$$J_1\left(\frac{4\pi B}{\lambda}\right) \approx \frac{1}{2}\left(\frac{4\pi B}{\lambda}\right) \qquad \left(B << \lambda/4\pi\right)$$

which simplifies the 1st harmonic to

$$v_I(t) \approx A\alpha_I \sin\left(\phi_e + \frac{4\pi r_0}{\lambda}\right)\left(\frac{2\pi}{\lambda}\right) B \sin(2\pi(f_r)t)$$

or, since $\tilde{r}(t) = B\sin(2\pi(f_r)t)$ and

$$\psi = \phi_e + \frac{4\pi}{\lambda} r_0 + \pi$$

and

$$\sin(\psi - \pi) = -\sin(\psi),$$

we get

$$v_I(t) \approx -A\alpha_I \sin(\psi)\left(\frac{2\pi}{\lambda}\right)\tilde{r}(t) \qquad \textbf{(2.6)}$$

which is the same as Equation 2.5 above.

2.3 Effects of the system parameters on the ARV signal

The effect of the system parameters on the ARV signal (including amplitude, distortion strength, and relative size) are:

1. The A coefficient, where $A = RE_0^2$:
    - The R is the reflection coefficient (in amplitude) from the target for far field reflections, and the transmission coefficient for near field interactions. Thus the amplitude of the ARV is directly linearly related to the energy reflected from or transmitted through the target.
    - E_o is the initial amplitude of the emitted wave. Thus the amplitude of the ARV output will vary as the energy of the emitted wave.
    - In tissue, the conductivity will also play a role. The greater the conductivity, the greater the loss of energy due to absorption. This causes the amplitude of the ARV signal to decrease.

2. The channel gains α_I and α_Q:
    - The output of each modulation will vary linearly as its gain.

3. The sine or cosine of the sum of the round trip distance from the antenna to the target in wavelengths ($\frac{4\pi r_0}{\lambda}$) and the phase delay (if any) of the electronics on the reflected wave or the transmitted wave (ϕ_e):
    - These factors are added together and the sine or cosine taken to determine the magnitude and phase of the relevant harmonic. The value of ϕ_e is irrelevant, it is simply a constant phase shift. The harmonics and the **I** and **Q** modulations both alternate between sine and cosine. This ensures that if one of the modulations is badly distorted, the other will be significantly less distorted. The worst-case

is a combined total of $\pi/4 + n\pi/2$ (n = 1, 2, 3…), which results in equal amplitudes (and distortions) for both channels.

4. The Bessel functions of the first kind.
    - The magnitude of each harmonic depends directly on the corresponding Bessel function magnitude, which itself depends on the relative amplitude of the target motion B. The Bessel function of the first kind, $J_N\left(\frac{4\pi B}{\lambda}\right)$, are shown in Figure 2.2 for B ranging from 0 to λ.

      For a B greater that about $\lambda/6$, the second Bessel function amplitude is a significant fraction of the first, which can lead to noticeable distortions.

It is important to note that variables 1 and 2 are simply gain values that do not affect the shape of the ARV signal, but variables 3 and 4 can and will distort the shape and the energy of the signal – especially the shape.

To illustrate the causes of the distortions represented above, let's examine the lower half of Equation 2.3 with some annotations:

$$N^{th}\ \text{harmonic} = (-1)^{N+1} A\alpha_I \sin\left(\phi_e + \frac{4\pi r_0}{\lambda} + \frac{M\pi}{2}\right) J_N\left(\frac{4\pi B}{\lambda}\right) \sin\left(2\pi(Nf_r)t + \frac{M\pi}{2}\right)$$

Distortion due to distance to target (amplitude of mixed wave)

Distortion due to relative amplitude of target vs. wavelength

Signal of interest

There are two major sources of distortion in this equation. The first varies the amplitude of the harmonic based upon the distance to the target. Some distances yield large, clear returns, others smaller, distorted returns. If the argument of the first term is near a multiple of 0 or π, then the return of the N^{th} harmonic will be near zero. Note that *this effect is purely dependent upon the*

relative distance to the target, r_0/λ, and so can occur with any target distance or amplitude given the right circumstances. Again, this is the reason that both modulations (90 degrees out of phase with one another) are used – so that there is never a target distance where a useable signal is not available from the ARV.

The second distortion term above is caused by a large relative amplitude of motion of the target. This is because a sine wave can be considered linear only over a small portion of a wavelength. If we want to keep the amount of distortion low, we must keep the relative amplitude of the target low so that the argument of the Bessel function is small. If we are interested in the amplitude (the shape) of the wave, we set a limit of J_1/J_2 (fundamental to harmonic) of 10 or greater. If we zoom in on the lower end of Figure 2.2 and take the ratio of J_1/J_2, the results are shown in Figure 2.4. It is clear that for $J_1/J_2 = 10$ we need to keep the phase of the Bessel function at 0.03 wavelengths of the carrier wave or less. Of course, the wavelength of the carrier wave depends on the medium through which it is traveling, so what do we use for this distance?

For our 2.4 GHz system, the wavelength is 12.5 centimeters and thus we are limited to a magnitude of motion in air of about 12.5 * 0.03 = 3.75 mm to ensure that the distortions are minimized. In tissue, this will be 2-6 times less, or about 0.6 to 1.9 mm. This condition could become difficult to satisfy for some targets and is of paramount importance if the correct shape of the wave is desired.

If we are only interested in the energy, we can tolerate more amplitude distortion, as the energy is the square of the amplitude. We can thus lower the J_1/J_2 ratio to about 3.2, which will give us a ratio of about 10 in energy. This allows B amplitudes of 0.0935λ, or about 1.2 cm in air or 2-6 mm in tissue. This is much more likely to be satisfied. As a result, the relative energy of the target motion, not its shape, is more likely to be accurately depicted by the ARV.

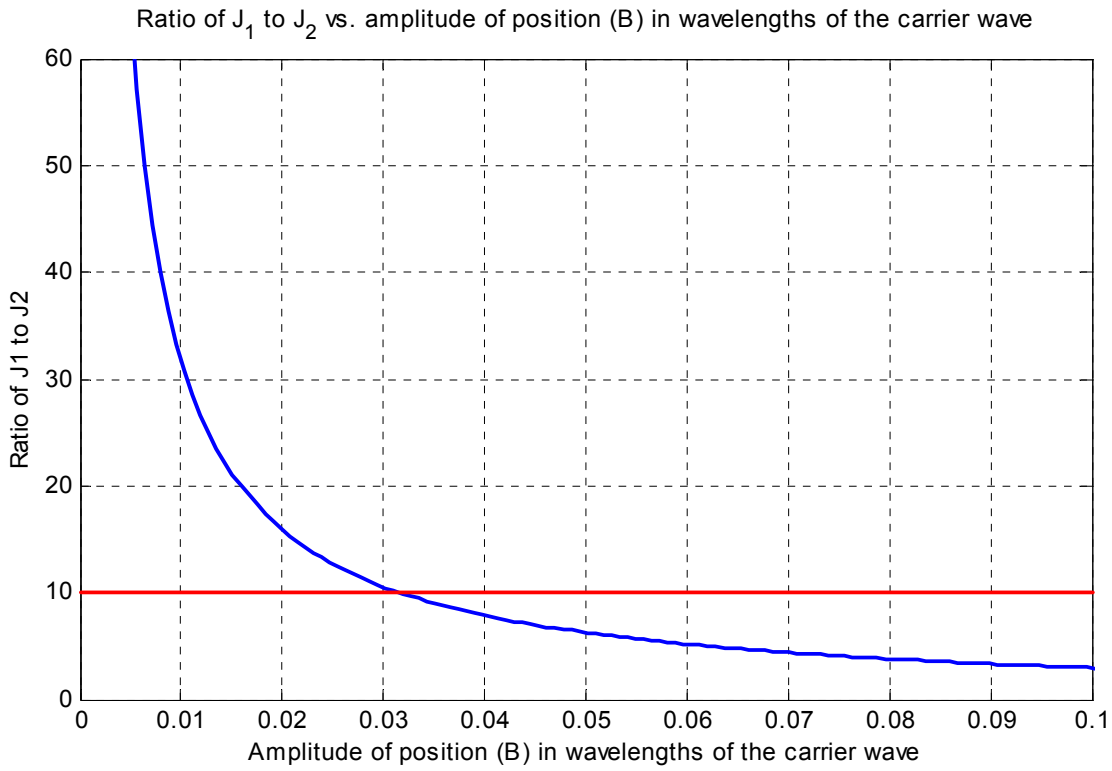

Figure 2.4. The ratio of J_1 to J_2. We want the amplitude of the first harmonic to be 10 times or more the energy of the second, so we should stay at or below a relative amplitude of 0.03λ, which corresponds to an amplitude of motion of 3.75 mm for the ARV.

Therefore, a researcher determined to measure the exact position of any tissue must satisfy the following constraints:

1. The amplitude of the motion of the target must be below about 0.03λ, in the medium in which the vibration is taking place.
    - Luckily, this is securely under the extent of motion for objects in the body that vibrate above 40 Hz or so, including the tracheal wall and the cheek. It may, however, be a problem for larger motions such as the tongue and jaw. In that case lower frequency sensors may be required in order to increase the wavelength and allowable motion amplitude. It should be clear that directly correlating motion to the ARV signal is difficult and should be approached with caution. If the distortions are

to be kept to a minimum, a periodic or quasi-periodic position target motion should be examined and the above restrictions followed. However, this does not mean that data gathered from non-periodic motion is not useful – it is just not likely that an accurate position signal can be derived from it. However, even distorted position information may be useful.

2. Both the **I** and **Q** channels must be acquired and examined for any signs of distortion due to harmonics. Normally, the channel with the largest amplitude is also the one with the least distortion and a simple energy calculation can be used to find the best signal. However, it is possible that the distorted channel signal could have an energy near to or even slightly greater than the clean channel signal, so an examination of the signal for distortion artifacts should be undertaken.
    - Once the antenna is in place and if it is not allowed to move with respect to the subject, the channels with the most distortion will normally not change.

2.4 What the ARV actually "sees"

The above derivations show that with proper care we may recover the original position signal, but the magnitude of the signal can vary significantly depending on several factors: *The magnitude and shape of the ARV signal depends on the distance from the antenna to the target (r_o), the amplitude of the emitted wave (E_o), the conductivity of the target (skin depth), the reflection/transmission coefficient (R), and (finally) the amplitude of the target vibration.* This magnitude can also be negative, resulting in a 180-degree phase difference for sinusoidal motion, but this does not affect the frequency content or location of transient events.

Therefore it can be said that under the right conditions, the sensor is able to detect the correct frequency content and phase (+-180 degrees) of the position of the target, but the absolute amplitude cannot be derived. The *relative* change in magnitude is correct, though, and with the above information can be very useful.

2.5 The near field

Any interaction of tissue with the near field can cause the same reaction as a far field reflection from a surface. The near field does not travel away from the antenna, it simply transits from one part of the antenna to another. Any disturbance, though, in this transmission, results in a change of the energy returned to the ARV and is detectable. For example, if a portion of tissue is in the near field, a certain amount of energy will pass through the tissue from one part of the antenna to another. To the ARV, this is indistinguishable from the reflection of transmitted energy from a target very near the antenna. In addition, the tissue in the near field significantly affects the input impedance of the antenna, changing the phase of the wave returning to the ARV. If the tissue is then removed from the near field, the amount of energy conveyed by the near field as well as the impedance of the antenna will change. In turn, this will change the phase of the returning wave, and this phase change due to the movement of the tissue in the near field is indistinguishable from that due to target movement in the far field. Therefore, near field interactions result in ARV signals that appear very similar to those derived from far field interactions and the above relationships are valid for tissue motion in the near field as well.

2.6 What about the Doppler shift?

Some explanations of the ARV signal include (and emphasize) the Doppler shift of the RF wave upon interaction with the vibrating surface. Indeed, there is always a shift in frequency when the target is moving toward or away from the

antenna. Let's look at the change in phase due to this Doppler shift. If our position signal is

$$r(t) = r_0 + B\sin(2\pi f_r t + \phi_r)$$

then the velocity is

$$v(t) = \frac{dr}{dt} = 2\pi f_r B\cos(2\pi f_r t + \phi_r).$$

The maximum value of the velocity would therefore be $\pm 2\pi f_r B$. For normal vibrations inside the trachea and cheek, the amplitude of vibration (B) is normally less than 1 mm. For argument's sake, let's say B is 3 mm and the frequency is 1000 Hz. Then the maximum velocity would be about 20 meters per second. The amount of Doppler frequency shift is given by

$$f_{SD} = \frac{2v}{c} f_0$$

so the maximum value of the shift is

$$f_{SD\_max} = \frac{40}{3x10^8} 2.4x10^9 = 320 \text{ Hz}$$

Since

$$\lambda_0 = \frac{c}{2.4x10^9} = 12.5 \text{ cm}$$

and

$$\lambda = \frac{c}{(2.4x10^9 + 320)} = 12.499998 \text{ cm}$$

which is a change of only 0.00001 radians in phase – very negligible compared to 0.15 radians, the change an the mixed wave due to the 3 mm motion. Therefore we may assume that the frequency is constant in our derivation.

2.7 Imbalance and the I and Q channels

The question has been raised as to what happens when α_I and α_Q are not equal. The gains in the ARV are designed to be equal, but a small difference is always possible. If this occurs the modulation is said to be unbalanced. However, as both the **I** and **Q** modulations (without distortions) represent only the relative motion of the position signal, it is inconsequential if **I** and **Q** have different gains. These gains will only affect the amplitude of each modulation and will not affect the frequency content of the signal. However, if the energy of the signal is used to determine the modulation with the fewest distortions, the gains should be examined in greater detail to ensure that they are roughly the same. A combinational approach is also possible and is discussed in Section 5.5.

2.8 Conclusion

It has been shown that the ARV sensor, operating as a phase modulation motion detector, can return an accurate position vs. time signal with correct frequency and phase but only relative amplitude. The amplitude of the ARV signal depends on the distance from the antenna to the target, the amplitude of the emitted wave, the conductivity of the target, the reflection/transmission properties of the target, and the amplitude of the target vibration. In addition, distortion of the signal can occur if the target is near a null of the mixed wave (distance distortion) or if the target amplitude of motion is large compared to a wavelength of the transmitted wave (relative amplitude distortion). In order to reduce the amount of distortion in the signal it is recommended to measure periodic or quasi-periodic motion with an amplitude of 3 mm or less. Near field interactions result in ARV signals that are similar to those cause by far field interactions and, since the antenna is used on-body, probably are the primary contributor to the ARV signal. The phase difference due to Doppler shifting can be neglected, and compensation for any unbalanced gain in the modulation circuitry is normally unnecessary.

3. Radiofrequency/tissue interaction

Human tissue interacts strongly with electromagnetic waves due to the tissue's relatively high permittivity ε and conductivity σ. The magnetic permeability μ, on the other hand, is relatively low and does not play a significant role in the interaction with electromagnetic waves. A significant amount of this section is based upon information found in the author's thesis [3].

There are three major modes of interaction between human tissue and EM energy. The first is reflection, where the EM energy simply reflects from tissue interfaces. The second is refraction, where the EM energy travels through the tissue and is redirected due to the shape and composition of the tissue. The third is absorption, where due to finite tissue conductivity the EM wave is absorbed by the tissue and the EM energy is transformed into heat. Interactions between the body and EM waves can be a combination of all three of these effects and are highly dependent on the geometry and location of the tissue.

3.1 Electromagnetic properties of human tissue

The electromagnetic properties of human tissue are not known to a high degree of precision. Most of the studies on the properties have been performed *in vitro* (out of the body) after death. There is very little data (for good reasons) on live human tissue, specifically skin, fat, cartilage, and smooth and skeletal muscle. Most of the values used in research come from canine or porcine (pig) tissue.

The permittivity describes how the material is affected by electric fields and is usually described as a dimensionless ratio in terms of ε_o, the permittivity of free space. This ratio of $\varepsilon/\varepsilon_0$ is termed the dielectric constant, an unfortunate name as it normally varies substantially with frequency and can therefore only be considered constant for a finite range of frequencies. A dielectric constant of 52 indicates the permittivity of the tissue is 52 times that of free space at the

frequency of interest. Since the displacement field D is equal to εE [6] and is continuous along dielectric boundaries, the strength of the electric field E inside the tissue would be 52 times less than the field outside if all electromagnetic energy transferred to the tissue (no reflections). This effect is due to polarization of the tissue, which opposes the electric field inside of a dielectric. Since the permittivity of the dielectric is always greater than that of free space, the electric field inside a dielectric is always decreased by the dielectric constant.

This process is analogous to the polarization of magnetic dipoles that occurs in magnetism, which is more easily visualized using a magnet and iron filings. The same effect takes place in dielectric materials placed in an electric field. Magnetic permeability, on the other hand, is very nearly that of a vacuum inside human tissue and is therefore not that important in a study of EM/tissue interaction.

Conductivity, however, is very important. Conductivity is the inverse of resistance and describes how much the electromagnetic wave is attenuated as it transits the material. Conductivity causes the absorption of electromagnetic radiation as free electrons and ions in the material are moved by the incoming wave's fields. The electrons and ions will seek the lowest energy state and rearrange themselves in such a way as to cancel out the incoming wave's field. Their success depends on how good the conductor is and whether resistance is present. The better the conductor, the faster and more complete this cancellation. In a "perfect" conductor, the electrons are infinitely fast and electromagnetic waves cannot penetrate at all. In a real conductor, the electrons and ions lose energy through collisions and the incident electromagnetic energy is dissipated and turned into heat. This is the process by which a microwave oven heats food, through the excitation and rapid vibration of water molecules. Thus electromagnetic waves lose energy is they travel through a conductive medium. A convenient measure of the rate of energy loss is the *skin depth*, usually expressed in centimeters, which is the depth into the medium that an

incoming wave penetrates before its amplitude drops to 1/e of its initial value (36.8%).

An example of how conductivity attenuates electromagnetic waves is the transmission of light through seawater. Seawater is not a good conductor (about a million times less conductive than most metals), but light is unable to penetrate more than about 30-40 meters. It is not due to sediments or suspended material (although that does make it worse), the electromagnetic waves are simply dissipated by the free electrons and ions in the water. In the process the water is heated. This effect is lessened in fresh water, so light can penetrate deep freshwater lakes (such as Lake Tahoe) to a much greater degree.

Keep in mind that conductivity is a function of frequency as well, but since we operate at one frequency for our purposes the conductivity is constant. The conductivity of human tissue in our frequency range of 1-3 GHz is about 1.1 to 2.5 S/m ([4], [5]), about 10-20 times less than sea water and about even with germanium, a semiconductor [6]. Still, the skin depths for skin and muscle are not insignificant and conductivity will play a role in the ARV/tissue interaction.

In order to choose valid approximations to the electromagnetic "constants", values were taken from [4], [5], and [7]. Agreement between sources is fairly good, with the difference in values attributable mostly to the moisture content and whether or not the measurement was in-vitro (out of the body) or in-vivo (in the

Freq in GHz	λ_o cm	Fat			Cartilage			Smooth muscle			Skin		
		ε_r	σ	d	ε_r	σ	d	ε_r	σ	d	ε_r	σ	d
1	30	15	0.2	10.4	22.5	0.12	21.0	56	1.4	2.8	33.5	0.7	4.5
2.5	12	13	0.4	10.3	21	0.13	21.0	52	2.2	1.8	33	1.0	3.1

Table 3.1. Dielectric constant, conductivity, and the resulting skin depth for biological tissues at approximately 1 and 2.5 GHz ([4], [5], [7]). ε_r is relative dielectric constant, σ is conductivity in S/m, and d is the skin depth in cm.

body) and whether the measurement was taken before or after death. Canine and porcine tissues are quite similar to human, and were used as an approximation in some cases so that before-death in-vivo measurements could be taken. The average results are shown in Table 3.1.

3.2 Plane-wave reflection from planar surfaces

It is illustrative to examine the simplest of all far-field interactions – plane wave scattering from planer surfaces. It is critical to understand that there must be a surface or dielectric discontinuity in order to reflect EM energy. A wave traveling in a homogenous medium will never spontaneously reflect; it will reflect only from surfaces of discontinuity. This model is simple because geometric effects, near field interactions, finite thickness of layers, interference between reflected and transmitted waves, and conductivity are neglected. Thus the wave is not attenuated as it travels through the dielectrics and all layers are considered planar with a thickness much larger than a wavelength. This will lead to reflection coefficients that are too high, as the layers are somewhat conductive and geometric effects always lower the effective cross-sectional reflective areas. Without conductivity, the electromagnetic property of interest is the permittivity, or in this case the *index of refraction* n:

$$n = \sqrt{\frac{\varepsilon\mu}{\varepsilon_o \mu_o}}$$

or, if we use ε_r (the ratio of tissue permittivity to free-space permittivity) and assume a permeability of $\mu = \mu_0$:

$$n = \sqrt{\varepsilon_r}$$

The percentage of energy reflected at each planar surface (the reflection coefficient R) is determined by the change in the index of refraction by the following simple formula [6]:

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2$$

where the wave is traveling from material 1 to material 2. The percentage of energy transmitted at each surface (transmission coefficient T) is simply 1 – R. For this example, the tissue layers in the neck were simulated. Figure 3.1 shows the relative values of the indices of refraction and the amount of reflection from each layer. The smooth muscle represents the walls of the trachea, with an air gap to represent either the trachea. At each interface the reflection coefficient is labeled.

It is evident that a good deal of the energy (54%) is reflected from the outer surface, making it difficult to get good penetration of far field waves into the body. The largest reflection, though, comes from the muscle/air interface at the front of the trachea and the air/muscle interface at the rear of the trachea. Transmission through either of these interfaces would result in significant reflections. The skin, fat, cartilage, and smooth muscle layers are not very different in n, so R is not large at those interfaces.

The thing to notice is that reflections can occur at any interface at which there is a discontinuity in the index of refraction. However, small discontinuities result in

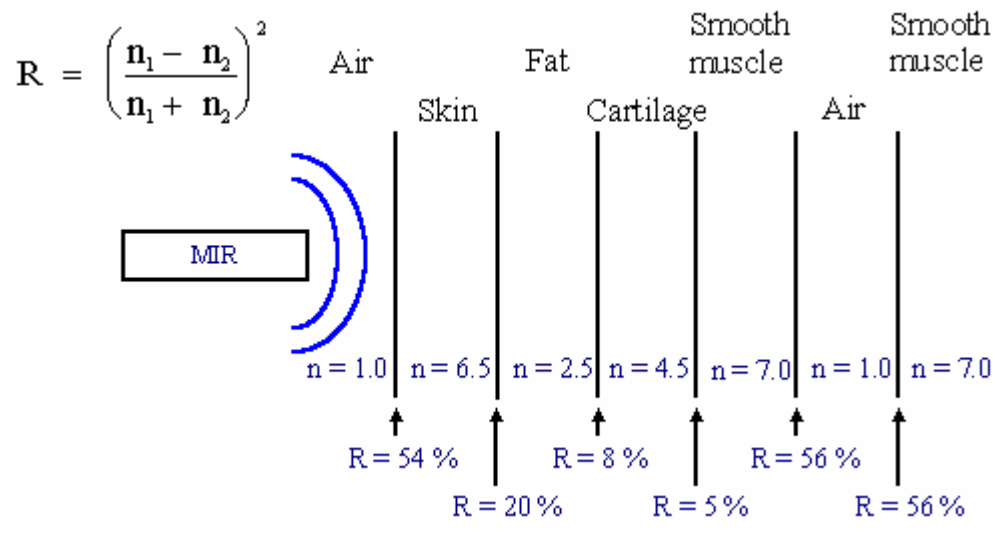

Figure 3.1. Simple planar calculation of the neck tissue layers' reflectivity, neglecting geometrical factors, multiple reflections, and conductivity. [3]

minor reflections, and the larger the indices of refraction the less is reflected. For example, there is an 8% reflection at the fat/cartilage interface, where there is a difference of 2 in the indices, but at the cartilage/smooth muscle interface there is only a 5% reflection, even though the difference is 2.5, due to the relatively large size of each index.

Please keep in mind, though, that this is a gross simplification of the entire process and is just included to get an idea of the structure and non-uniformity of the tracheal region, where the ARV is often used. It is very unlikely that the ARV is emitting planar waves which travel through the layers of tissue above – in fact, as we shall see, it is unlikely that the ARV, with its antenna on the skin, is emitting any traveling waves at all.

3.3 Reflection due to conductivity

As mentioned earlier, conductivity can also cause incident EM energy to be reflected. This is because the incoming EM wave causes the charge carriers in the medium to vibrate in concert with the EM wave, dissipating the EM energy as heat. However, this vibration of change carriers also acts as a source of EM radiation, so that the EM energy is re-emitted and appears to bounce off of the medium. For perfect conductors, the reflection is also perfect, and the EM energy completely reflects from the conductive surface. For nonperfect conductors, there is always some penetration and dissipation of EM energy. The skin depth can be thought of as a metric of the amount of energy loss that electromagnetic waves suffer in conductive materials. It is the depth in the material at which the amplitude of the electromagnetic wave has been reduced by 1/e. The formula given by Griffiths ([6], p. 370) for skin depth d is:

$$d = \frac{1}{k_{-}}$$

where

$$k_{-} = \omega\sqrt{\frac{\varepsilon\mu}{2}}\left[\sqrt{1+\left(\frac{\sigma}{\varepsilon\omega}\right)^{2}}-1\right]^{1/2}$$

Using the values for smooth muscle at 2.5 GHz given above (which have to be transformed from relative to absolute permittivities), the approximate skin depth is 1.8 cm. This means that for every 1.8 cm in smooth muscle tissue the wave has to travel, it loses about 63% of its amplitude (about 9 dB down in energy) to the conductivity of the tissue. The skin depth is greater for other tissues (such as fat and cartilage) because of their lower conductivity. Calculated skin depths for various tissues at 1 and 2.5 GHz are given in Table 3.1.

For near fields, the skin depth will not be a significant issue. The thickness of all of the tissue between the skin and the front wall of the trachea is on the order of a centimeter, so even with a round trip the conductivity is not likely to dissipate more than half of the energy of the wave. For far fields, though, any transit through 3-4 cm (required simply to construct the far field) would constitute a significant reduction in the energy of the return.

For example, for waves traveling along the length of the folds, the skin depth could very easily be exceeded and the wave could lose 18 or more dB of its power due to conductivity losses alone. Also, waves traveling into the trachea from the side or back of the neck could encounter several centimeters of tissue and suffer large energy losses as well. Therefore it is always a good idea to position the antenna so that the EM waves travel through as little tissue as possible to the target(s) of interest.

3.4 Diffraction

Diffraction occurs when waves bend around obstacles in their path rather than scatter from them. A wave is much more likely to diffract than reflect if the wavelength of the wave is about the same as or greater than the cross-sectional

dimension of the object. A classic example is the circumlocution of a sea buoy by a long-wavelength sea wave. When confronted with the buoy, the wave simply wraps around the buoy and continues on its way. Very little energy is reflected in this interaction. Another example is the far-field transmission of energy around the vocal folds. In the smooth muscle of the folds, the wavelength of the ARV transmitted wave is about 12.5/7 = 1.8 cm, far greater than the 3-4 mm width of the glottis. Thus the vast majority of the ARV energy would simply diffract around the glottis. This is for traveling waves, though, and near field interactions can lead to other effects, which we discuss below.

3.5 Antenna/skin interactions

Our tissue/EM interaction discussion has so far focused on the traveling far field waves, which propagate from an antenna through a medium. This is perhaps not surprising, since the far fields are much easier to describe and are the conventional way to discuss antenna emissions. However, it is very important to remember that not all fields produced by an antenna radiate. Some simply begin on one part of the antenna and terminate on another. These non-traveling waves are termed the “near field”, because they exist mostly near the antenna. They are usually ignored in antenna theory and design, as antennae are normally designed to move RF energy from one location to another, and the structure of the localized near fields is not relevant to that task. However, the near fields transform into the far fields, and are thus absolutely essential to the construction of far fields.

Another reason that the near fields are not considered in antenna design is that whereas the far fields are relatively simple and uniform, the near fields are complex and vary significantly from one type of antenna to another. However, since we are using the ARV to study vibrations located well within the near-field region of the antenna, it is of paramount importance that we include them in our discussion.

For a conventional antenna operating at 2.4 GHz, the near field extends about 12.5 centimeters in air, but is only strong very near (within a few centimeters of) the antenna. However, the placement of lossy, non-homogeneous dielectric human tissue directly against the antenna significantly changes the way the antenna operates. The tissue does not simply shorten the wavelength of the transmitted electromagnetic wave; it changes everything regarding the performance of the antenna.

3.5.1 Near vs. far field interactions

The dielectric of the body can significantly disrupt the normal operation of an antenna. In order to form far-field (traveling) waves, the near fields of the antenna must join up at a distance at or greater than a wavelength. *This means that interaction between the tissue and the antenna inside that distance will be necessarily via the near field.* Since the near field becomes the far field, any interaction with non-uniform lossy dielectrics can significantly decrease or even completely destroy the ability of an antenna to create far field radiation.

For example, dielectric lenses are often used to shape the electric fields of antennae so that the directionality of the transmitted waves is affected. However, these lenses are composed of a homogenous dielectric, normally thin compared to a wavelength, and placed uniformly over the antenna. If such a dielectric lens is located over only part of the antenna, it can completely disrupt the formation of propagating waves. Thus a non-uniform dielectric presence in front of an antenna can cause the antenna to cease transmitting energy in the far field, and the only interaction possible will be with the near field. This non-uniformity can exist anywhere in the near field region ($r < \lambda$) around the antenna, but is most disruptive when it is placed on or very near the antenna, as when the ARV antenna is placed on the skin.

To illustrate the differences between the near and far fields, Figure 3.2 contains a diagram of the near and far fields for a dipole. Note that near the dipole, the electric field lines begin on the positive change and end on the negative charge. This is typical of the near fields – they don't propagate away from the dipole, they simply oscillate in place. The far fields, on the other hand, do not form from the near fields until (in this example) a distance of about three wavelengths. Thus interactions very near the antenna will be with the near field only and significantly disrupt the formation of far-field traveling waves. Several excellent animations regarding radiation and the attendant transformation of near fields into far fields can be found at http://web.mit.edu/jbelcher/www/anim.html#Radiation.

Finally, the lossy dielectric tissue is polarized by the near field and can increase the reach of the near field. The tissue "pulls" in the near field lines, extending their reach. As an example, we will discuss below an experiment where a finger held between the edge of the antenna and the skin can facilitate the detection of

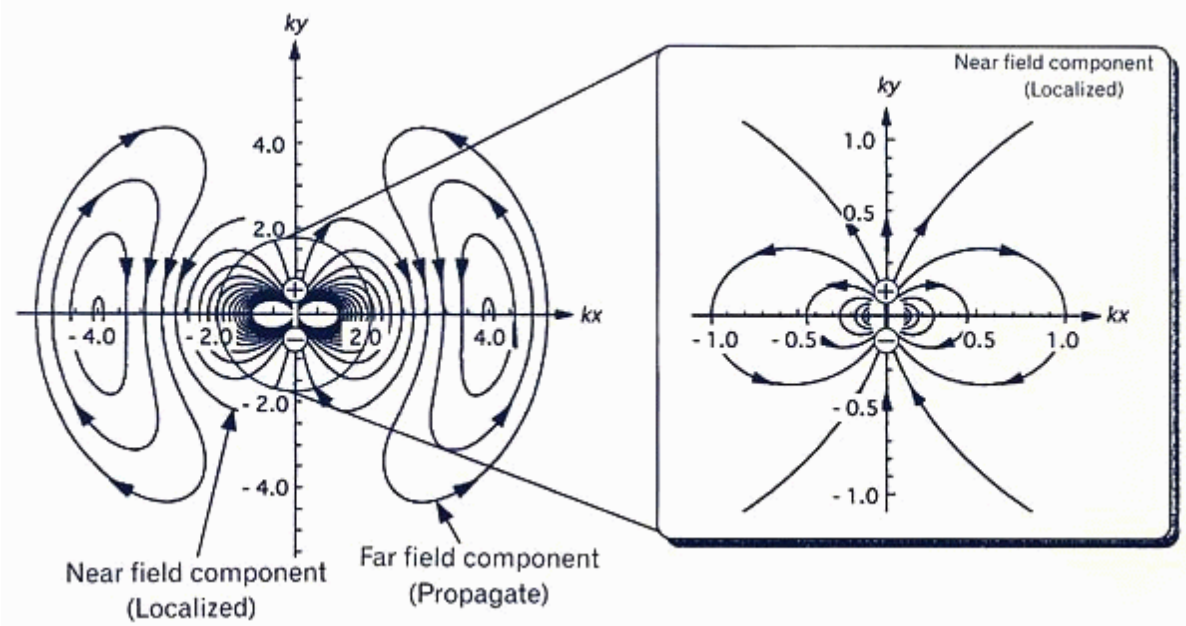

Figure 3.2. Near and far fields for a dipole. The near fields oscillate in place, they do not propagate as do the far fields. However, any disturbance in the near field oscillation will also be detected by the ARV. From here.

fold closure. Fingers held on the edges of the antenna can greatly increase the amount of near field penetration compared to holding the antenna against the neck with a non-dielectric substance, such as an elastic band.

3.5.2 Near-field interaction with the ARV

Now that we have determined that the ARV interacts predominately through the near field, it is useful to understand how the movement of tissue inside the near field interacts with the ARV.

Every antenna has a characteristic input impedance. The generator of the ARV "sees" this impedance and the transmitted and reflected wave are affected by it. It is conventionally designated by

$$Z_A = R_A + jX_A$$

where R_A is the resistance of the antenna and X_A is the reactance. The resistance can be further broken down into the radiation resistance R_R and the loss resistance R_L, and the reactance in terms of the capacitance (neglecting the inductance of the antenna, which is normally small compared to the capacitance):

$$Z_A = (R_R + R_L) - \frac{j}{wC}$$

The loss resistance is a function of the materials with which the antenna is constructed and is not changed by the dielectric interaction. However, the radiation resistance does change significantly due to the proximity of the lossy dielectric – it is significantly reduced.

The capacitance of the antenna is greatly increased by the presence of the dielectric. The increase is directly proportional to the relative dielectric constant of the dielectric, or in this case a multiple of 5-7 times the original capacitance, depending on how much of the near-field is filled with dielectric. This means that the reactance of the antenna is also significantly reduced.

Since the magnitude of the impedance is

$$\left|X_A\right| = \sqrt{R_A^2 + X_A^2}$$

and the phase is

$$\angle X_A = \arctan\left(\frac{X_A}{R_A}\right),$$

we can examine the effects of the dielectric. It tends to make X_A somewhat smaller, and R_A much smaller. Thus the magnitude of the impedance of the antenna is significantly smaller, and the phase of the impedance is significantly different. This impedance change due to the presence of the dielectric will result in a change of phase of the reflected wave. Therefore, movement of a non-uniform dielectric in the non-uniform near field of the antenna will change the amount of dielectric present in the near field, changing the effective impedance of the antenna, resulting in a detectable signal from the ARV.

Since both R_A and X_A decrease significantly in the presence of the lossy dielectric, instead of free space, the electromagnetic energy in the antenna "sees" a virtual ground outside the antenna. This causes the energy to almost completely reflect from the antenna/skin interface. This would seem to indicate that there could be no interaction, since almost all energy is reflected, but remember that this reflection can contain near-field interactions since they do not radiate away from the antenna. Thus the change of impedance due to the dielectric presence in the near field is still detectable. Far field emissions, on the other hand, are not likely due to the almost complete reflection of energy at the virtual ground.

3.5.3 Examples of near field interaction

In this section we will examine three experiments designed to demostrate the detection of motion inside the human body using only the near field. This is possible by purposely designing or modifying an antenna so that the far field is minimized. One way to do this is to coat the front of the ARV antenna with foil to

disrupt any potential far-field activity. The first experiment uses copper foil placed on the ARV antenna to disable the far-field activity that may be taking place, although the ARV antenna is naturally a very poor far-field antenna in air at 2.4 GHz. The second shows how tissue can be used to pull in the near-field of the antenna. Either of these experiments can be easily confirmed using an ARV and one of the rectangular, planar antennae. The third uses an antenna designed to have virtually no far field at all and is also easily reproducible.

In the first, we cover the antenna from corner to corner with adhesive copper tape (Figure 3.3). This conductor, placed so near the antenna, significantly disrupts any far-field emissions that may be present. At the very least, the far field performance will be significantly degraded. The near field, though, is not significantly affected on the portions of the antenna that are not covered with foil. When placed on the trachea, the waveform observed (both at and below the vocal folds) *is identical to the waveforms observed without copper tape on the antenna*. This indicates that either the far field is not contributing at all to the signal, or the far field contribution is virtually identical to the near field. The latter would seem to be much less likely than the former - an excellent signal is possible using the near field alone, so the presence of a far-field-based signal is not required. Other configurations of copper (including covering all but 2 mm of one side) yield similar results, but the corner configuration is simplest to repeat.

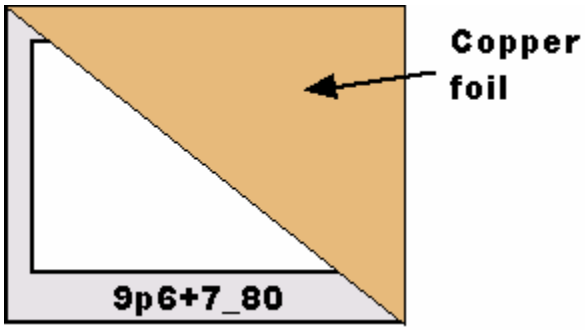

Figure 3.3. Copper foil used to destroy any far-field emissions from the ARV antenna.

The second experiment involves placing the ARV antenna (without foil) about 1 cm from the skin in front of the trachea by the vocal folds. No signal is observed in this position, which would indicate that neither the far or near field is contributing. A finger is then placed between the edge of the antenna and the

laryngeal prominence (Figure 3.4). The finger, in this position, should also significantly degrade any far field emitted by the antenna as it is only present over a small area of the antenna and would significantly disrupt any far field formation. The near field, on the other hand, is "pulled" into the finger and its reach extended due to the polarization of the finger. Its reach is significantly improved by the presence of the dielectric. In this position, a weak but significant vocal fold signal is observed. Since the far field is not present, the near field must be the primary means of interaction with the tissue. The near field uses the finger as a conduit and is able to extend into the trachea and interact with the vocal folds.

The last experiment uses a simple dipole antenna designed to have almost no far field, but a significant and relatively simple near field. Each arm of the dipole is about 2.5 cm long, and the arms are oriented parallel to one another, with a spacing of about 1 cm (Figure 3.5, the dimensions are not critical). When this

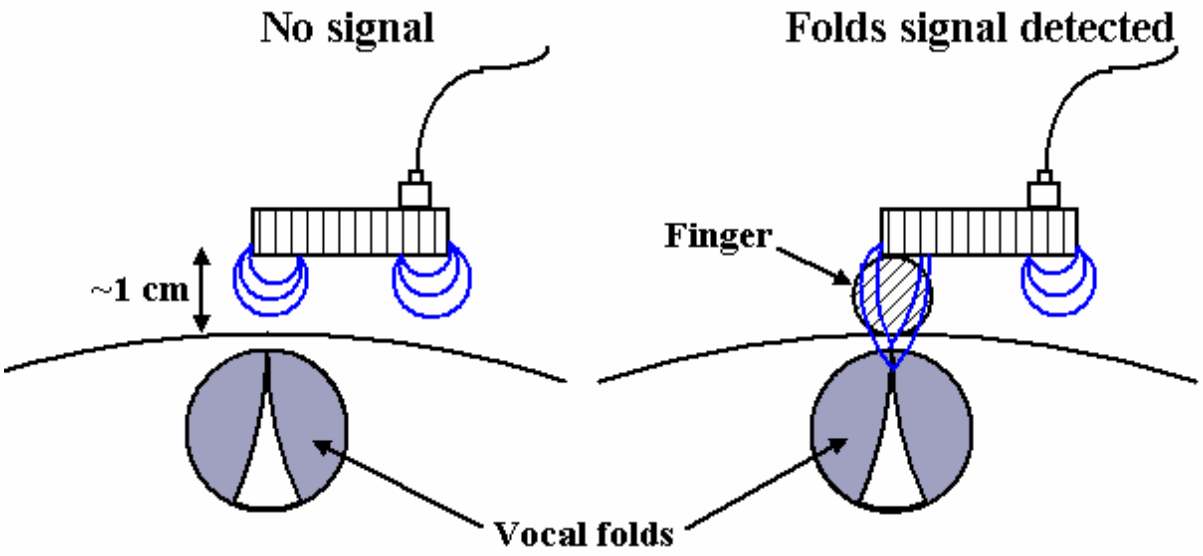

Figure 3.4. Illustration of a simple example of near-field interaction with the vocal folds. The antenna is held about 1 cm away from the laryngeal prominence (LP) and registers absolutely no signal when speech is produced by the subject (left). However, when a finger is placed between the side edge of the antenna (right) and the LP, a weak "folds closing" signal is observed. The finger would disrupt any far-field activity, but would facilitate the near-field interaction through polarization of the finger tissue. The near field gets "pulled into" the tissue and its reach is extended through the polarization field of the finger.

antenna is placed lengthwise on the trachea, a signal very similar to that observed with a conventional antenna is observed. Thus, this near-field antenna is perfectly capable of reproducing the results obtained with the conventional ARV antenna, indicating that both operate in the near field.

These experiments show clearly that the far field of the antenna does not contribute significantly to the output of the ARV. In the first, any far-field output would be virtually eliminated by the presence of the foil, and the signal observed is virtually identical to that observed when the foil is not present. In the second, the presence of the finger is very unlikely to affect the far field in a positive manner, but is able to channel the near field into the trachea and back. And in the third, an antenna with very limited far field capability and very simple near field returns a signal virtually identical to that obtained using a conventional ARV antenna.

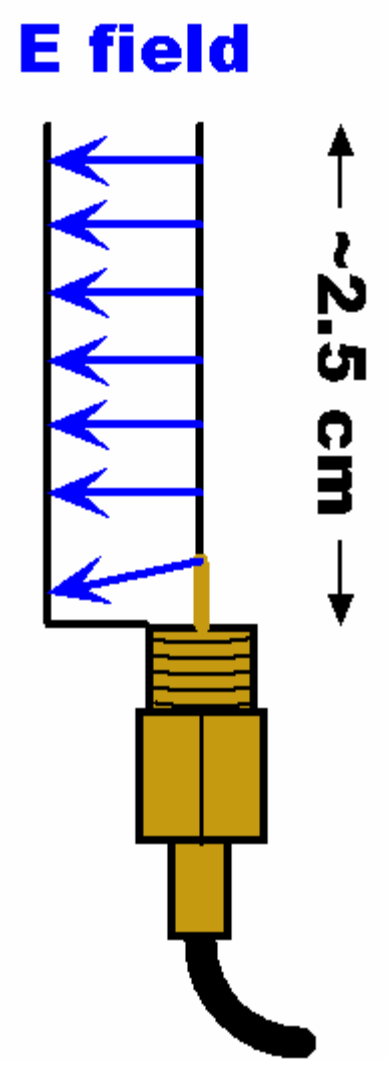

Figure 3.5. Near-field antenna constructed with a female SMA connector and two 2.5 cm wires, spaced about 1 cm apart. There should be little or no far field, but significant near field between the wires.

3.6 ARV antenna near-field model

The ARV antenna is a simple patch antenna, and therefore its near field should be very similar to that of a dielectric filled capacitor, with fringing fields located at each edge leading from the upper surface to the lower. It is these fringing fields that give rise to far-field traveling waves.

Nonetheless, it is instructive to be able to see what the fields look like for our antenna. Therefore, a model of one of Aliph's skin antennae was constructed

using WinFeko, a EM finite-element simulation tool. The model was as accurate as possible, emulating the patch size, thickness, and shape as well as the antenna feed point. Near fields from different planes are shown in Figure 3.6, located at the end of this Section. It is clear that near the edges of the antenna, there are significant near fields extending 4-5 mm from the antenna (white circle).

3.7 Analysis of trachea & vocal fold ARV signal

As discussed above, the near field of the antenna is more than sufficient to account for the observed waveform. In addition, the near field of the ARV is capable of measuring subcutaneous tissue vibration, as evidenced by the strong signals returned at the cheek, nose, and lower trachea, far away from the vocal folds. However, when placed near the laryngeal prominence of most subjects, a sharp feature is observed during normal voiced speech (Figure 3.7). It was shown ([3], p.114) that this feature corresponded exactly to the closing of the vocal folds during voiced speech, and that it was in the same location as fold closure as measured by an electroglottograph (EGG), which measures vocal fold contact area (although there were distinct differences, such as during breathy speech, shown in Figure 3.8, which is from the same dataset as Figure 3.7). It was originally thought that this feature was due to a “water hammer” effect from

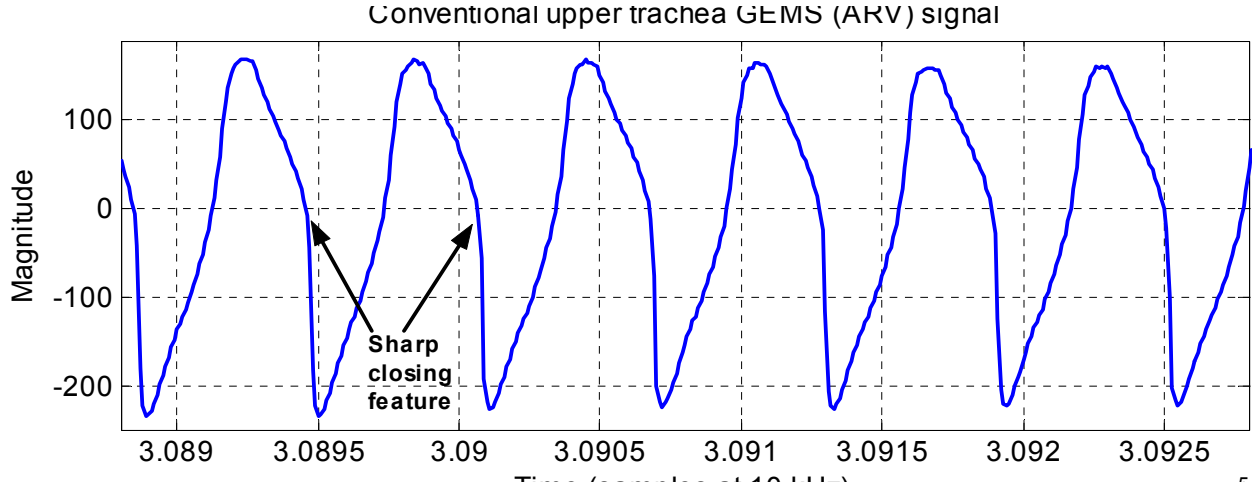

Figure 3.7. Conventional upper tracheal GEMS (ARV) signal with sharp fold-closing feature.

fold closure, but now it seems that it is likely due to the near field interaction with the vocal folds, as discussed above. The incomplete conclusions contained in prior ARV/tissue interpretations arose from the failure to include the possibility of near-field interactions.

The near field of the ARV antenna is not uniform, and even with the tissue acting as a conduit to the near field, it is not likely that the near field from the antenna can penetrate more than a centimeter or so into the tissue, at it drops in strength by about 20 dB going from 1 mm to 5 mm in Figure 3.6. Adding the lossy effect of the conductive tissue, it is likely that only the anterior tips of the folds will contribute to the overall tracheal signal. This explains why vocal fold nodules

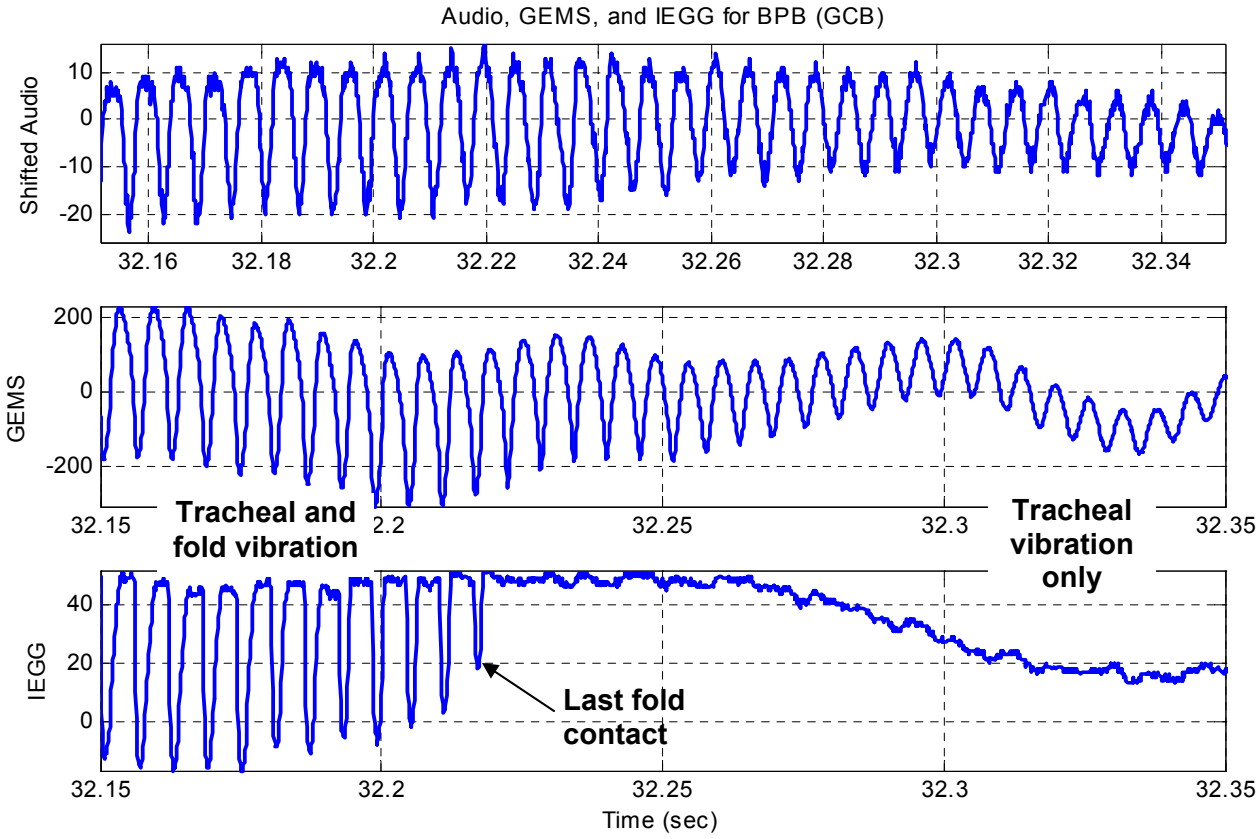

Figure 3.8. Shifted (to account for the much lower speed of sound) acoustic data (top), corresponding GEMS (ARV) data (middle), and inverse EGG data (bottom) for cessation of breathy speech. The GEMS was used for this experiment, but similar results have been noted using the ARV. Note when the EGG stops recording data that the sharp feature in the GEMS slowly disappears and is completely gone by 32.3 seconds. Data from [3], p.128.

were not detected using the ARV in a previous experiment ([3], pp.115-116). Those nodules were located about 1.5 cm from the front of the subject's neck, and were not detected at all by the ARV, even though their influence on the folds' operation was profound. In addition, for true falsetto speech only the outer cover of the vocal folds is vibrating and there is little contact between the folds [8, 9]. For many subjects during falsetto speech the ARV indicates no contact, just a vibratory signal. As an example, data from a male subject is shown in Figure 3.9. The transformation from normal (chest) speech to falsetto at about the same pitch is illustrated. It is clear that a significant amount of the signal is lost when

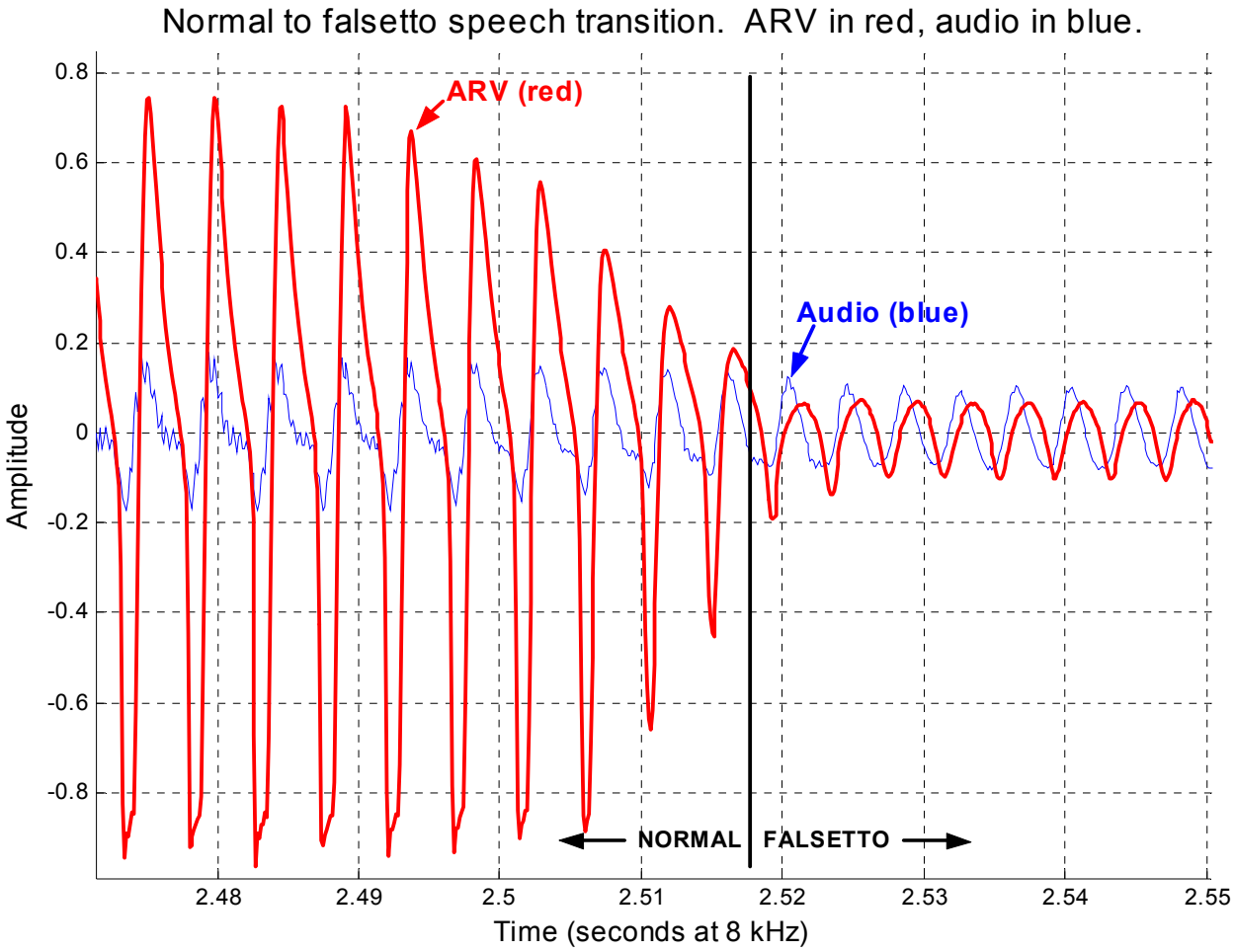

Figure 3.9. Normal to falsetto speech transition for a 35-year-old native English-speaking male. ARV is red, recorded audio (at the mouth) is blue. Note the loss of amplitude (about –18 dB), indicating that for this subject with the antenna in this location the signal for normal speech is mostly vocal fold contact. Also, note the change in the character of the audio as the folds separate – there are no longer many overtones present in the signal.

falsetto speech is produced, in this case about –18 dB, indicating that for this subject with the antenna at this particular location the ARV signal was mostly vocal fold contact in nature. This behavior has been observed for many subjects, with the highest loss in amplitude for the normal-to-falsetto transition observed for the subjects with the least amount of tissue between the trachea and skin. It can be difficult to get useable data since the volume and pitch should remain relatively constant during the transition. This data again suggests that the ARV signal taken near the laryngeal prominence (LP) is a combination of tracheal vibration and vocal fold contact area.

Further evidence that the LP signal is a combination of vibratory and vocal fold contact can be derived from a study of the tracheal signal below the folds and at the LP. Data from a male speaker is shown in Figure 3.10, with the lower

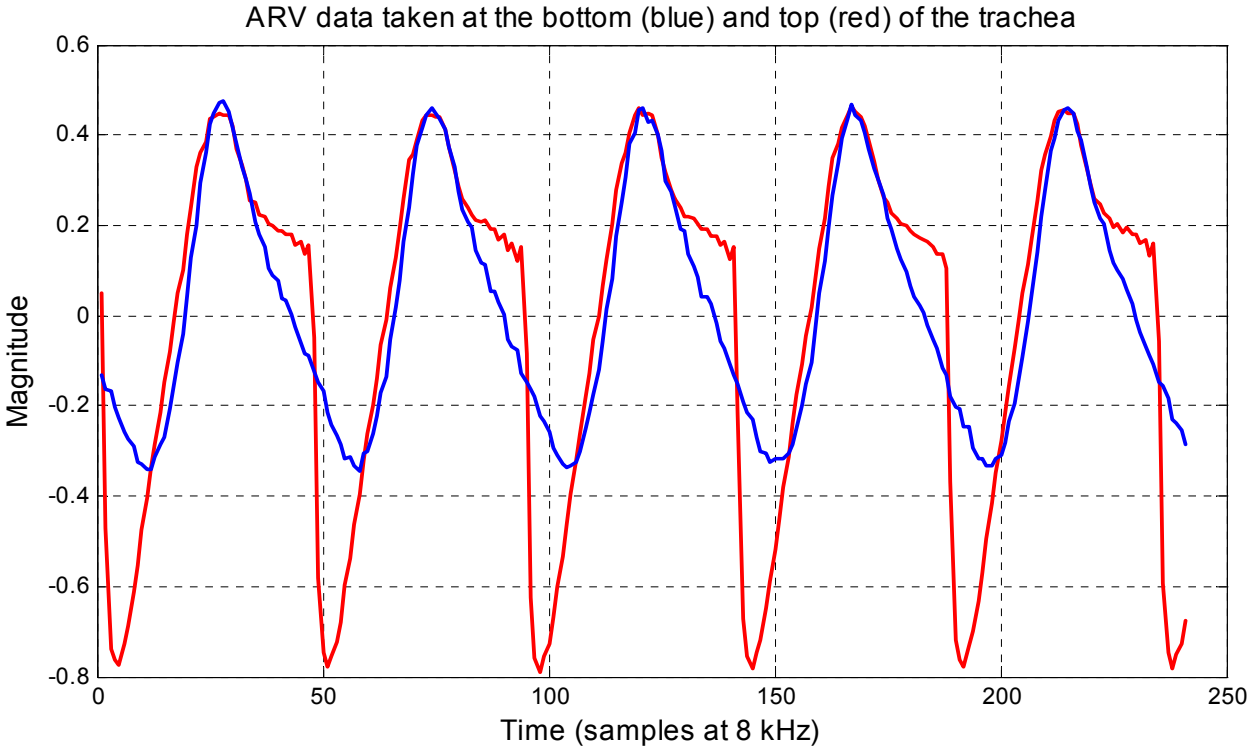

Figure 3.10. ARV signals taken from the lower (blue) and upper trachea (red) at the same pitch. Note the distinct lack of a sharp closing feature in the data from the lower trachea, although the magnitude and shape are similar to the data from the upper trachea. The data from the lower trachea is simply tracheal vibration, and the data from the upper trachea a combination of tracheal vibration and anterior fold closure. (ARV1_trimmed.wav)

tracheal antenna position shown in blue and the LP position in red. The lower tracheal position is several centimeters away from the folds and a fold component to its signal is very unlikely. Therefore the lower tracheal signal is composed tracheal vibration alone. The LP position, however, may contain both tracheal vibration and vocal fold contact signals. The pitch and vocal effort were approximately the same with the antenna in both positions. It is clear that the proximity to the folds significantly increases the sharp feature that denotes fold closure and that a sizeable portion of the ARV signal near the folds is directly related to anterior fold closure. It is also clear that just a few centimeters away from the folds, there is no fold contact area in the signal and only vibratory motion is being detected by the ARV. Near the LP, the vocal fold contact signal is simply superposed onto the tracheal vibration signal.

To illustrate this superposition of vibratory and vocal fold contact signal, data was taken from a three-channel experiment in which audio, GEMS (previous to ARV, but similar), and electroglottograph (EGG, which is proportional to vocal fold contact area) were recorded simultaneously during breathy speech (Figure 3.8). The GEMS and EGG data in Figure 3.8 were used to form a signal composed of both (Figure 3.11). Since this data is of the end of breathy speech, only the anterior portion of the folds is still in contact. Thus the EGG signal represents the vocal fold contact area of only the anterior portion of the folds. The ARV (GEMS) signal has almost lost the closing feature and is therefore reverted almost entirely to a vibrational signal. If we add the EGG to the vibrational ARV, we get a signal (bottom plot in Figure 3.11) that looks very similar to the normal ARV signal shown in Figure 3.7, which is from the same data. This is a strong argument that the normal tracheal ARV signal is a combination of tracheal vibration and anterior vocal fold contact area. The combination of the symmetric vibrational data (symmetric because it is a driven oscillator) with the asymmetric anterior vocal fold contact area signal (asymmetric because current only flows during anterior fold contact; the magnitude of the signal has been reversed in this example) results in the characteristic asymmetric ARV signal with its sharp closing feature.

It must be stressed that probably only a small portion of the anterior folds affects the signal due to the lack of a contact signal for patients with vocal fold nodules. The contact signal is not caused by any sort of waveguide effect from the closure of the folds nor is it due to a reflection of any energy from the glottis or from the rear wall of the trachea, it is simply a manifestation of the anterior fold motion's interaction with the near field of the antenna, which changes the input impedance of the antenna and thus the phase of the returning wave.

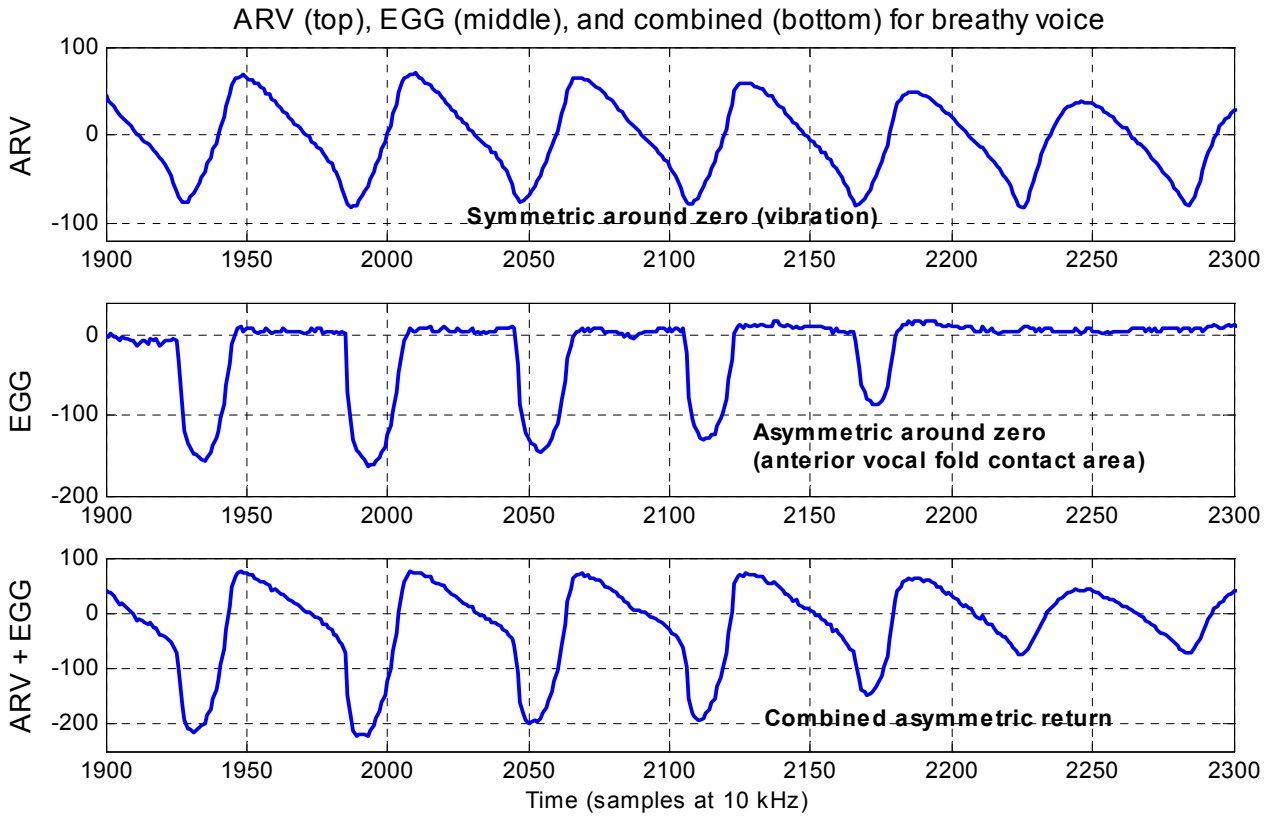

Figure 3.11. Demonstration of the combination of tracheal vibration and vocal fold contact area to form the conventional ARV tracheal/folds signal. This data was taken that shown in Figure 3.8, from about 32.19 to 32.23 seconds, during the gradual cessation of breathy speech, when the vocal folds are pulling apart and the only contact is in the anterior portion of the folds. The top plot is the ARV (GEMS), which has lost most of the contact feature and is primarily vibratory in nature. The middle plot is the EGG signal, which in this case represents the anterior vocal fold contact area, since the folds are gradually separating and the last part to separate is the anterior portion. The last plot is the first two added together. It is clear that the combination of the vibratory and anterior vocal fold contact area signal results in a signal that looks exactly like the normal ARV signal. It has the correct shape and is also asymmetric, as it should be. The vibrational signal is symmetric around a zero point because it is a driven oscillator (there is a slight drop due to low frequency noise in this example), but the anterior vocal fold contact area is asymmetric (a signal only passes when there is contact; the magnitude has been reversed here). Thus the superposition is asymmetric as well, as we observe with the normal ARV signal shown in Figure 3.7.

It must also be emphasized that this superposition of vibratory and anterior fold contact area signals will vary in strength depending on how much of the near field energy is able to penetrate the skin and interact with the folds. In some subjects with a significant amount of tissue around their trachea, it will be difficult to get the needed penetration and the signal will look much like the top plot in Figure 3.11. On the other hand, some subjects with prominent trachea will have very large contact area signals. *The weighting of each signal in the superposition will depend on the subject's tissue structure, the antenna, and the position of the antenna.* For most male subjects during normal (not falsetto) speech, the weighting is about even, as shown in Figure 3.10. For many female subjects, the weighting may be more vibrational than contact area, as there is typically more tissue between the skin and the vocal folds, reducing the amount of near field that is able to penetrate into the folds.

This theory certainly answers some questions as to the operation of the ARV:

1. *Why are the ARV and EGG similar when the folds are in contact and very different when they are not?* Because the ARV also detects the oscillatory movement of the trachea and the EGG only measures vocal fold contact area, which can be a large portion of the ARV signal.

2. *Why does the ARV not detect the folds well on those with recessed trachea or significant amounts of tissue in front of the trachea?* Because the near field cannot penetrate deeply enough to detect the folds.

3. *Why does the presence of a hand holding onto the ARV antenna usually improve the ARV signal?* Because the fingers of the hand can help channel the near field into the trachea and vocal fold region.

4. *Why is it so much more difficult to get a good signal on the cheek?* Because the cheek does not vibrate as much as the trachea, and its inner surface is

much farther away (~10 mm compared to 2-3 mm), so the strongest part of the near field is located in a dielectric that is not significantly changing.

So what does the composition of the ARV signal near the vocal folds mean in terms of using the ARV signal in speech processing applications? Not much, really. The two parts of the signal are intimately related as the opening and closing of the folds rapidly varies the subglottal pressure and forces the trachea to vibrate at the same frequency as the folds. The sharp fold closure signal simply adds more harmonics to the signal, increasing its usefulness, as the tracheal vibration signal does not contain a significant amount of harmonics above 1 kHz due to strong damping by the tissue.

3.8 Superposition

It cannot be overemphasized that the ARV will detect any targets that move or interact in any way with the fields of its antenna. If there is more than one object moving, the resulting ARV signal will be a superposition of all movement signals detected. In addition, as described above, some components of the signal may suffer from overtone distortion and the signal near the vocal folds is a superposition of tracheal vibration and vocal fold contact. All of these factors can make the "meaning" of the ARV signal very difficult to determine. Care must be exercised not to falsely attribute the ARV signal to a single source when two or more may be contributing to the signal. Instead of the "meaning" of the signal, it is often more useful to examine the signal obtained for useful information and use that information to enable meaningful applications.

3.9 Conclusion

It has been shown that the interaction of EM energy from the ARV with human tissue comprises two significant parts, both of which are due to interactions with the non-traveling near field. Vibrational returns can be collected from various

places around the body such as the nose, cheek, and trachea. Near the vocal folds, the near field can often penetrate enough to interact with the folds and return a vocal fold contact signal similar to an EGG. For tracheal locations near the folds, the ARV signal is a superposition of tracheal vibration and anterior vocal fold contact area, resulting in a sharp "folds closing" feature that is not available elsewhere on the neck or head (Figure 3.12). This sharp closing feature greatly increases the number of harmonics in the ARV signal, increasing its value for use in speech applications.

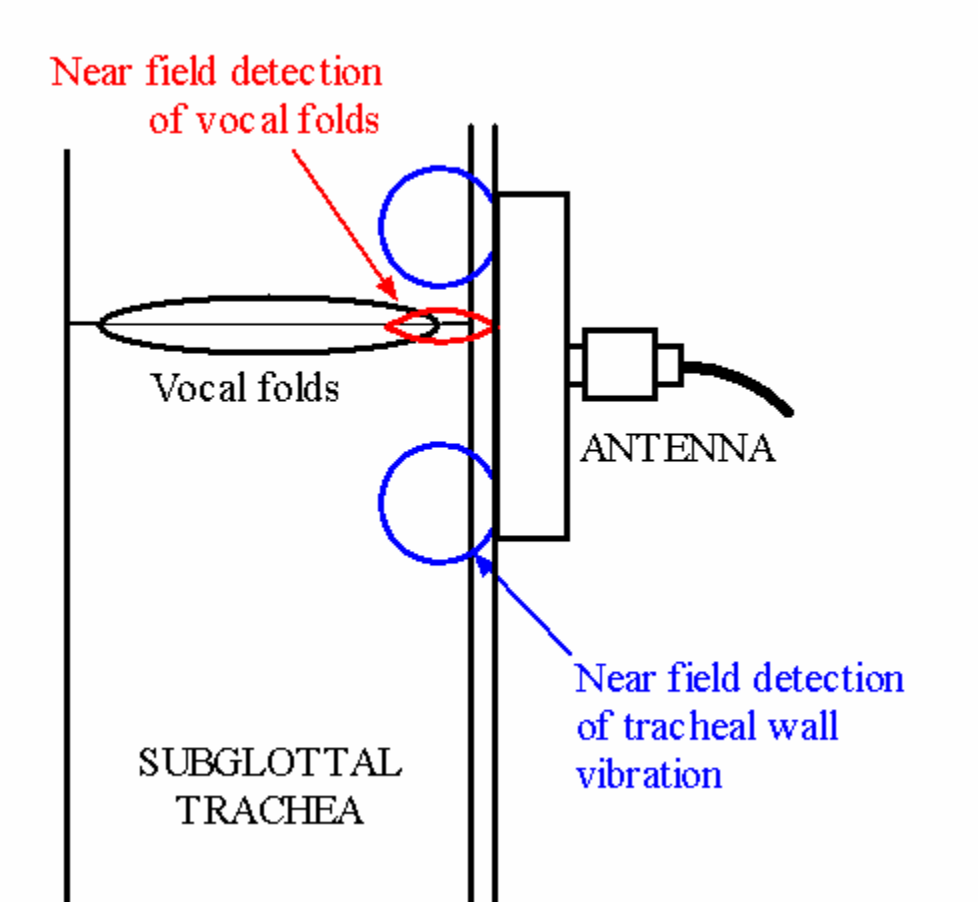

Figure 3.12. Model of antenna near field interaction with trachea and vocal folds. Since vibration is easily detected away from the folds, but closing signal only seen near the folds, the antenna must detect both tracheal vibration and anterior fold motion through the near field of the antenna. The vocal folds are detected with horizontally oriented fields that travel through them similar to an EGG.

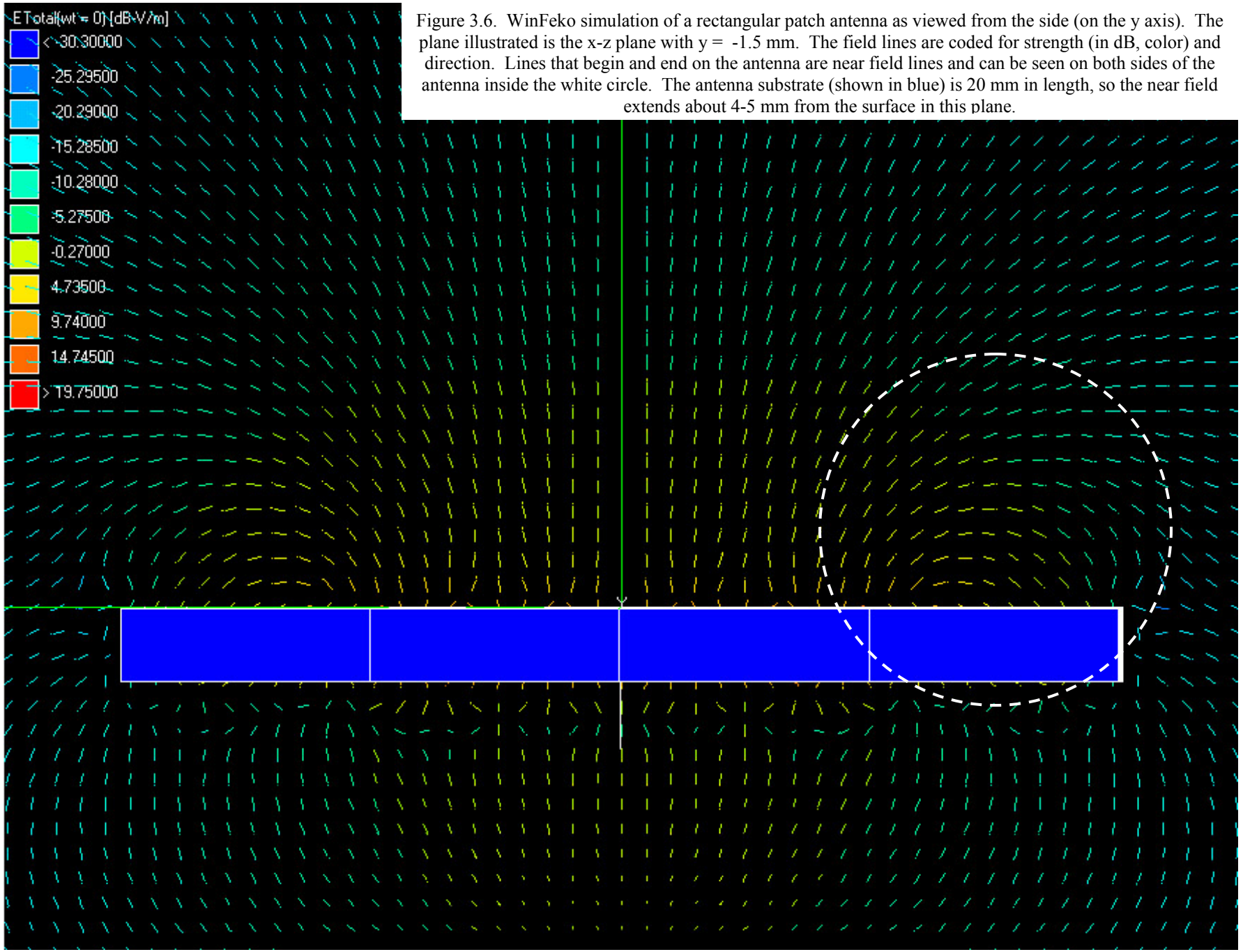

Figure 3.6. WinFeko simulation of a rectangular patch antenna as viewed from the side (on the y axis). The plane illustrated is the x-z plane with y = -1.5 mm. The field lines are coded for strength (in dB, color) and direction. Lines that begin and end on the antenna are near field lines and can be seen on both sides of the antenna inside the white circle. The antenna substrate (shown in blue) is 20 mm in length, so the near field extends about 4-5 mm from the surface in this plane.

4. Proper use of the ARV

In order to use the ARV effectively, the researcher must understand the basics of how the ARV operates as well as how the tissues vibrate during speech. Depending on where the ARV is used, different information can be extracted from its signal.

4.1 Physical components of the ARV

The ARV consists of three major assemblies:

1. The circuit board and case
2. The cable from the case to the antenna
3. The antenna

All three assemblies must be in good working order and be properly applied before the ARV will function at full capacity. It is especially critical that the cable be in good condition (no bends or breaks in the insulating shield) and that the antenna be positioned so that the near field (located near the edges) will be able to penetrate as deeply as possible. For experiments when a person cannot hold the antenna against the skin, it should be flush with the skin to ensure repeatability. It is critical that the antenna be held firmly against the skin, but this may be uncomfortable for some subjects. If measurements are being taken in air, the antenna must be positioned so that a significant return from the object under study is obtained. There are other necessary items such as the battery and its charger, but their use is covered in the user's manual that comes with the ARV unit.

4.1.1 The circuit board and case

The ARV is a continuous wave phase modulated radiofrequency device that operates at a frequency of precisely 2.45 GHz. It uses a monostatic antenna configuration with reverse polarity SMA connectors (which are required for FCC

compliance). The maximum theoretical RF power output is 1 mW while the effective power delivered to the target depends on the efficiency of the antenna. It returns a voltage signal, limited to +- 2.5 V, which indicates the amplitude of motion of a periodic or quasi-periodic motion of a target in its field of view. It is capable of measuring very small (micron order) vibrations from 20 Hz to 8 kHz both inside and outside the body.

The ARV works by continuously emitting a radio wave and allowing the reflected wave to mix with the emitted wave (in phase, or I modulation) and the emitted wave shifted by 90 degrees (quadrature, or Q modulation). This mixing results in two signals that complement each other in terms of sensitivity and distortion. That is, when one modulation returns a small, highly distorted signal, the other will return a large, clear signal (see Section 2 for more details). This is the reason that both **I** and **Q** modulations are output from the ARV, for the amplitude and distortion of one varies as a sine wave and the other as a cosine wave. With both modulations, the variation in amplitude for an oscillating target is only +-0.7 to 1.0 of its nominal value, as opposed to +-0.0 to 1.0 with a single modulation.

For a target with oscillatory motion, the signal is a mix of the fundamental vibration frequency and its overtones. The presence of the overtones can lead to distortions in the signal for large (> 3 mm peak-peak) motion signals. Occasionally a pitch doubling can be seen in the ARV signal, but this is normally a sign of poor placement and/or relative amplitude (Bessel) distortions, as discussed in Section 2. In the former instance the antenna placement should be verified that it is flat against the skin, and in the latter the other channel output should be used.

The ARV is mounted in a metal case to help reduce RF noise interference and is powered with a small rechargeable battery. For more details, please see the user's manual.

4.1.2 The cable

The cable is important in two ways: First, to comply with FCC regulations, the connector on the ARV board is a reverse polarity SMA cable. Therefore you will need an adapter or a reverse-polarity cable (supplied with the ARV) in order to connect antennae to the ARV. Second, the cable is the primary source of spurious ARV signals and the primary path for RF interference to enter the ARV.

The cable is somewhat flexible to allow the position of the antenna to be easily varied. However, if it is flexed beyond about 45 degrees, especially near the ends, the insulation can become damaged, allowing RF energy to escape and enter through the break in the insulation. This is the cause for the great majority of noisy ARV signals. It is absolutely imperative that the cable be well supported, especially near its ends, so that it cannot bend too much and damage its outer conductive layer. If it is damaged, it can cause a significant amount of noise to appear in the ARV signal and the cable must be replaced.

4.1.3 The antenna

The antenna is the critical component of the ARV system as it allows the RF energy to interact with targets in air or in the body through its near field. A poorly placed or performing antenna can cause the return of the ARV to be poor or nonexistent.

Unlike normal transmissive antennae, the ARV antenna is placed very close the targets with which it interacts. As shown in Section 3, this means that the near field is the primary means of target/ARV interaction. The distances *r* from the antenna associated with the near field is:

$$\text{Near field:} \qquad r < \lambda$$

where λ is the wavelength of the wave in the medium – about 12.5 cm in air.

Interactions with antennae in free space are well understood – there are many 2.4 GHz antennae available that are very efficient at getting the RF energy into and out of the air. Propagation through the air is a relatively simple problem because air is homogeneous and its dielectric and magnetic properties are very nearly those of vacuum. However, when the antenna is placed near the body, which is heterogeneous, has a high dielectric constant, and is conductive, the problem becomes much more difficult. The wavelength of an electromagnetic wave is directly proportional to the index of refraction of the dielectric through which it travels, but what is the index of refraction in the body? Skin, fat, cartilage, muscle, blood, are all present in varying quantities and all have different indices of refraction, ranging from about 2 to more than 6. In addition, all of this tissue is well within the near field of the antenna, where the fields are much more complex. Luckily, the presence of this heterogeneous, conductive dielectric prevents far-field emissions from forming – so strictly speaking, there is no wavelength. The antenna can only interact with the tissue via the near fields, which originate on one antenna element and terminate on another and do not propagate like intermediate and far fields. Finally, the nonzero conduction of the tissues of the body lead to skin depths (the distance in the tissue where the energy of the wave has been reduced to 1/e of the original) of only a few centimeters. This hinders deep penetration of the ARV energy into the body.

The present antennae supplied with the ARV are the second generation of antennae developed at Aliph and work well for most subjects. They were designed to be efficient in air at about 3.3 GHz, which drops to about 2.4 GHz when placed near the skin. This efficiency is defined by the amount of RF energy transmitted into the skin – the more energy transmitted from the antenna, the higher the efficiency. Since little or no far-field generation is expected when the antenna is placed on-skin, the observed shift in transmission must be from increased interaction of the near field with the body. Since the near field does not radiate away from the antenna, an increase in the energy transmitted by the

antenna through the near field would not seem likely. However, the conductivity of the tissue causes the transiting near field to lose energy, simulating a transmission of energy away from the antenna. Energy is still being lost by the antenna, but not through transmission – it is being lost through near-field conductive losses.

This shift in resonance frequency due to the interaction between the near field of the antenna and the skin is not easily predictable nor well understood, and an antenna that works well for some will not work for others due to differences in tissue dielectric. The near field is strongest near the surface of the antenna, though, so thin layers of dielectric (such as cellophane tape) can significantly affect the resonance properties of the antenna. Therefore the antennae supplied with the ARV can be used with up to three layers of cellophane tape to help tune the resonance of the antenna when it is placed upon the skin of the subject.

The ARV user's manual goes into great detail about how to find the best antenna configuration for your subject. Normally, any of the three antennae will result in a useable signal from the subject, but one will usually work better than the others. Since we can reduce the skin shift using cellophane tape, the idea is to start with the antenna that has the lowest resonant frequency in air (labeled 9p6+8_80) and see if the shift due to the subjects skin is small enough to return a large signal. If not, the 9p6+7_80 and 9p8+6_80 antenna should be used, and failing that cellophane tape can be added to the 9p6+6_80 antenna to further reduce the skin shift. This technique has been used successfully with many subjects, but can be time consuming. In some cases, a single layer of 3M duct tape has also resulted in excellent matching of the antenna and subject.

4.2 In general

For good results when not being held on the subject by hand, the antenna MUST be placed flat on the skin. If it is not, the signal can vary significantly in amplitude

and composition. This is one of the most difficult aspects to using the ARV – constructing an antenna mount that can hold the antenna in place for a variety of users and positions. We have found that an adjustable elastic neckband works relatively well for tracheal studies. Even better if the antenna can be held on the subject by hand, but this isn't always practical.

4.3 From the neck

If placed just below the laryngeal prominence (LP, the notch in the cartilage at the top of the "Adam's apple"), the ARV signal will be a combination of fold closure and tracheal vibration, and will look similar to the top plot in Figure 4.1 (located at the end of this Section). The closing and opening points are shown and can be highlighted by taking the derivative or just the difference (bottom plot in Figure 4.1), which is not the same as the derivative but is sufficient to highlight the vocal fold closing and opening times. The closing times are best used to calculate pitch and for pitch-synchronous processing, where the size of the processing window during voiced speech is determined by the pitch of the speech. Also, the ARV signal can be used simply to determine that voiced speech is being generated, which can be useful in acoustically noisy environments.

Finally, the ARV signal can be used to generate an excitation function. The excitation function is the acoustic driver of the vocal tract resonating system and contains no phonetic information. It is similar to the carrier frequency for a radio wave. An excitation function can be used in a source-filter model [8] to calculate the transfer function of the vocal tract, which can in turn be used to determine the phoneme that is being spoken. The excitation function for the vocal tract should contain all the frequencies needed to generate speech – that is, the fundamental frequency (the pitch) and enough overtones to sufficiently excite the vocal tract. *It is not necessary for the excitation function to exactly represent either the glottal flow or glottal pressure in order to be effectively used as an excitation function.*

The important characteristics for an excitation function are that they are both time-aligned with the speech being produced and contain the same frequencies as the produced speech. The ARV signal is perfectly time-aligned with the speech and, depending on the strength, can have overtones up to 4 kHz present, which is sufficient to excite the vocal tract for most applications.

As the ARV operates at the speed of light and the speech propagates at the speed of sound, a simple delay on the ARV signal brings it and the speech into alignment. The delay is calculated as the distance from the glottis to the microphone recording the sound divided by the speed of sound and is normally on the order of around 0.7 msec. This small delay is easily compensated for, but can usually be ignored as it has very little effect on the processing of 10-20 msec windows.

As demonstrated in Section 3, the tracheal/LP ARV signal is related directly to subglottal pressure and vocal fold movement and as such is not subject to phonetic variation. That is, like the excitation function, it is the same for all phonemes. The only information it contains is the pitch and location of the overtones.

4.4 From the cheek

The signal from the cheek reflects the amount of energy resonating inside the oral cavity. For example, when an “ah” is produced, most of the energy is radiated out of the mouth and very little resonates inside the oral cavity. This causes the cheek ARV signal to be weak for “ah”. By contrast, for “ee” there is a significant amount of energy resonating in the oral cavity. As a result, unlike the tracheal signal, the cheek signal is modulated with regard to phonetic content. As such, it may be able to assist in the identification of phonetic content in high acoustic noise environments. The ARV is not, however, normally sensitive

enough to detect the extremely small vibrations due to unvoiced speech production.

4.5 From the nose

A significant signal may be detected on the nose. Even more than the cheek signal, the nose signal is modulated with respect to phonetic content and may be used to identify nasal phonemes. It is also not normally sensitive enough to detect vibrations associated with unvoiced speech production.

4.6 Conclusion

The ARV can be used to detect vibrations inside and outside of the human body. The quality of its signal is sensitive to damaged cables, the antenna match to the skin, and the position of the antenna on the body. However, anywhere you can feel vibration on your body, it is possible to get a useable ARV signal (assuming that it is above 20 Hz and below about 8 kHz). It may require re-matching the antenna or using a different antenna altogether, but with proper care, any vibration inside or outside the body can be detected.

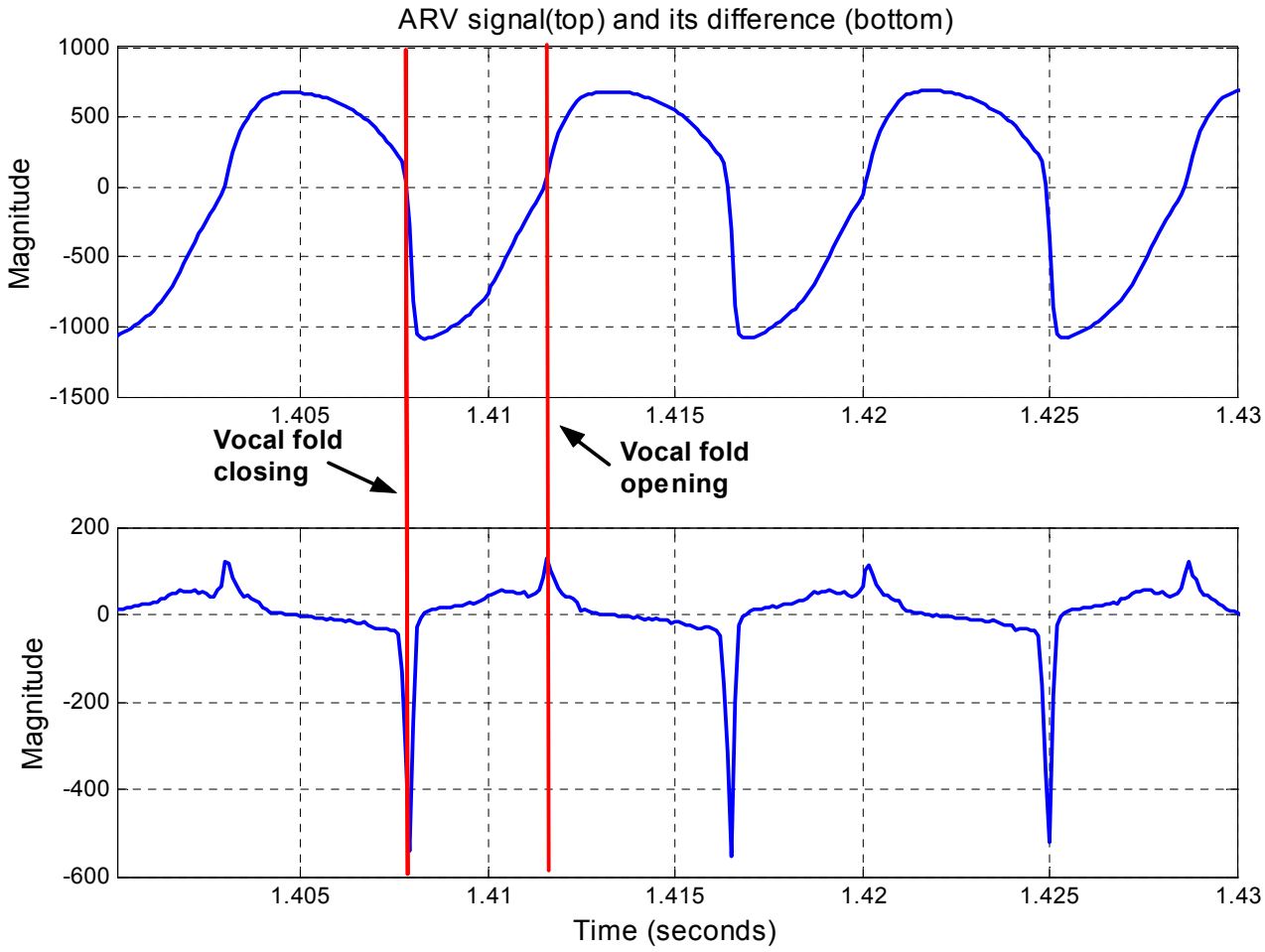

Figure 4.1. Standard ARV signal (top) and its difference (bottom) with closing and opening times marked. The difference is an effective, simple tool for locating the opening and closing points. The closing peak is much larger than the opening, as the closure is more rapid.

5. Generating speech features using the ARV signal

It is relatively simple to generate useful speech features from the ARV signal, depending on where the signal is gathered. These features may be used in a variety of applications.

5.1 Pitch

If the ARV is used on the trachea near the vocal folds, a sharp feature associated with their closure can be obtained. For very accurate, noise-immune pitch, the zero crossings of these sharp features can be interpolated using a simple linear interpolation technique. This can result in pitch estimates with accuracies greater than 0.1 Hz but requires a clean ARV signal that contains a strong fold closure signal.

For vibrations captured in other areas such as the lower trachea, cheek or nose or for less demanding applications where accuracy on the order of a few Hz is sufficient, any conventional acoustic pitch technique (such as autocorrelation or cepstral algorithms) may be used to generate the pitch data. By necessity a voicing on/off detection algorithm must be employed, and it has been found that a simple energy threshold based algorithm is sufficient for most applications.

Using these methods, a very accurate pitch obtainable every single glottal cycle is easily realized. This allows very close study of vibrato, jitter, and other pitch-based phenomena, even in heavy noise environments.

5.2 Voiced Excitation Function

One of the most useful and unusual data streams derivable from the ARV is the voiced excitation function (VEF), which can be used in a source-filter model of the vocal tract to characterize and synthesize speech [8]. The VEF can take several forms, from simple pulse trains to sophisticated wave shapes, but MUST

BE TIME-ALIGNED WITH THE SPEECH AND CONTAIN THE SAME FREQUENCIES AS THE SPEECH in order to function correctly as an VEF. If the produced speech has frequencies not contained in the VEF, the calculated transfer function could be unstable. Preferably, the VEF would precede the recorded acoustic waveform, as delay in a filter is very easy to model but advance is not. Many synthetic VEF do not align well with the produced speech, leading to inaccurate transfer function estimates. It is necessary for the VEF to exactly align with the produced speech at every time sample if accuracy and naturalness are to be maintained.

5.2.1 Pulsed excitation

The pulsed excitation has the advantages of being simple, robust, full-bandwidth, and easy to code. It is, however, probably the poorest approximation to the physical VEF, but that is not germane to many applications.

The pulsed excitation consists of both opening and closing pulses, which vary in amplitude and time. The closing pulses are stronger and take place over a shorter time than the opening pulses. Both are important if maximum accuracy is desired, but (depending on the speaker) the opening pulses may sometimes be dropped with little effect on accuracy. Closing points can normally be easily located using the tracheal ARV signal (see Figure 4.1).

The pulses may be located using the same zero-crossing interpolation technique as with the pitch above, but a simpler way to find the pulse points is to take the difference of the ARV signal and locate the peaks (see bottom of Figure 4.1). The closing peak is usually quite large, but the opening peak can be small or nonexistent. There are many common peak-finding algorithms that can be used for this application.

Once the peaks are located, the pulses are located at the peak points. The closing pulses are normally negative and the opening positive, to correlate with

the supraglottal pressure. The relative sizes of the opening and closing pulses is normally about 1/3, but more natural speech is possible if the size of the peak is used to determine the size of the pulse, as shown in Figure 5.1. When using the difference method, any abrupt change in the ARV signal can cause a peak to appear, but normally these peaks are small compared to the closing and opening peaks. These peaks may represent real phenomena (such as multiple vocal fold openings) or they may be due to a noisy return. Care must be taken in their interpretation, and a threshold should be used so that small peaks are ignored.

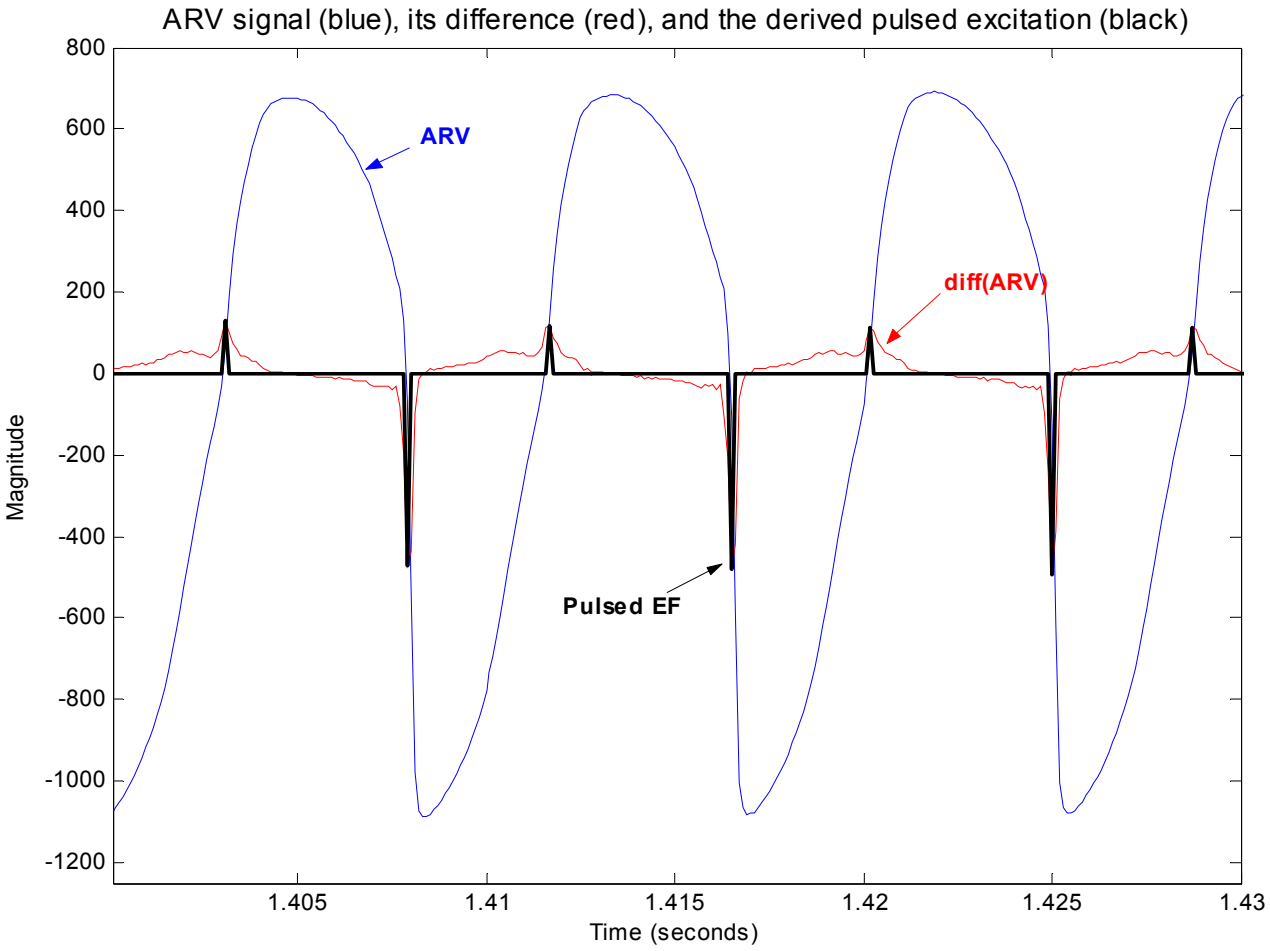

Figure 5.1. The ARV signal (blue), its difference (red), and a simple pulsed excitation function (black) that places the closing and opening points at the peaks of the cycle.

It is possible, and even desirable, to use only the closing peak pulses as the excitation function. The closing peaks are easier to locate and result in a flat-spectrum full-bandwidth excitation function that is simple to code and corresponds exactly to the frequencies found in the audio (Figure 5.2). This makes the closing-peak pulsed excitation function an excellent source in a source-filter model.

5.2.2 ARV signal as the excitation

It is possible to use the ARV signal itself as an excitation function. It is perfectly time-aligned with the voiced speech that is produced and requires no calculations or interpolations that might slightly change the time relationship between the ARV

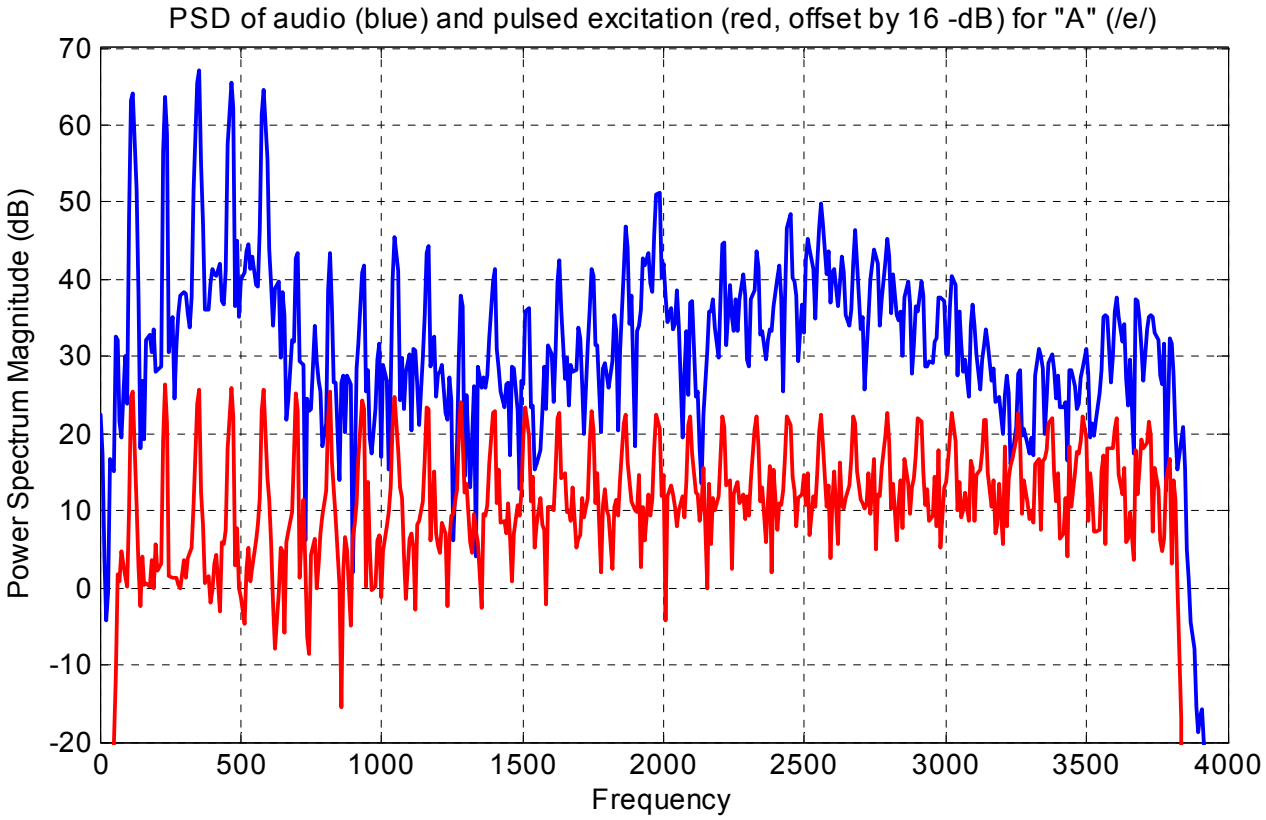

Figure 5.2. PSD of the audio (blue) and the pulsed excitation (red, offset by –16 dB) for "A" (/e/). Only the closing peaks were used to construct this excitation. Note the full bandwidth of the PE and the exact correlation with the frequency content of the audio. The PE has no formant information and will make an excellent excitation function for a source-filter model of the vocal tract.

and speech. After all, in the methods described above, the pulse locations will only be accurate to about +- ½ sample, which could introduce some inaccuracy in a subsequent calculation of the transfer function. As the ARV signal is time-aligned with the movement of the folds and the subsequent voiced speech, it contains the same frequencies the produced speech contains in the ARV's spectrum (see Figure 5.3). Thus the ARV signal can be used to derive a transfer function for the vocal tract with good accuracy. The energy of the ARV above 3 kHz, though, is low, so a pulsed excitation may be more appropriate to ensure adequate excitation of the entire spectrum of interest.

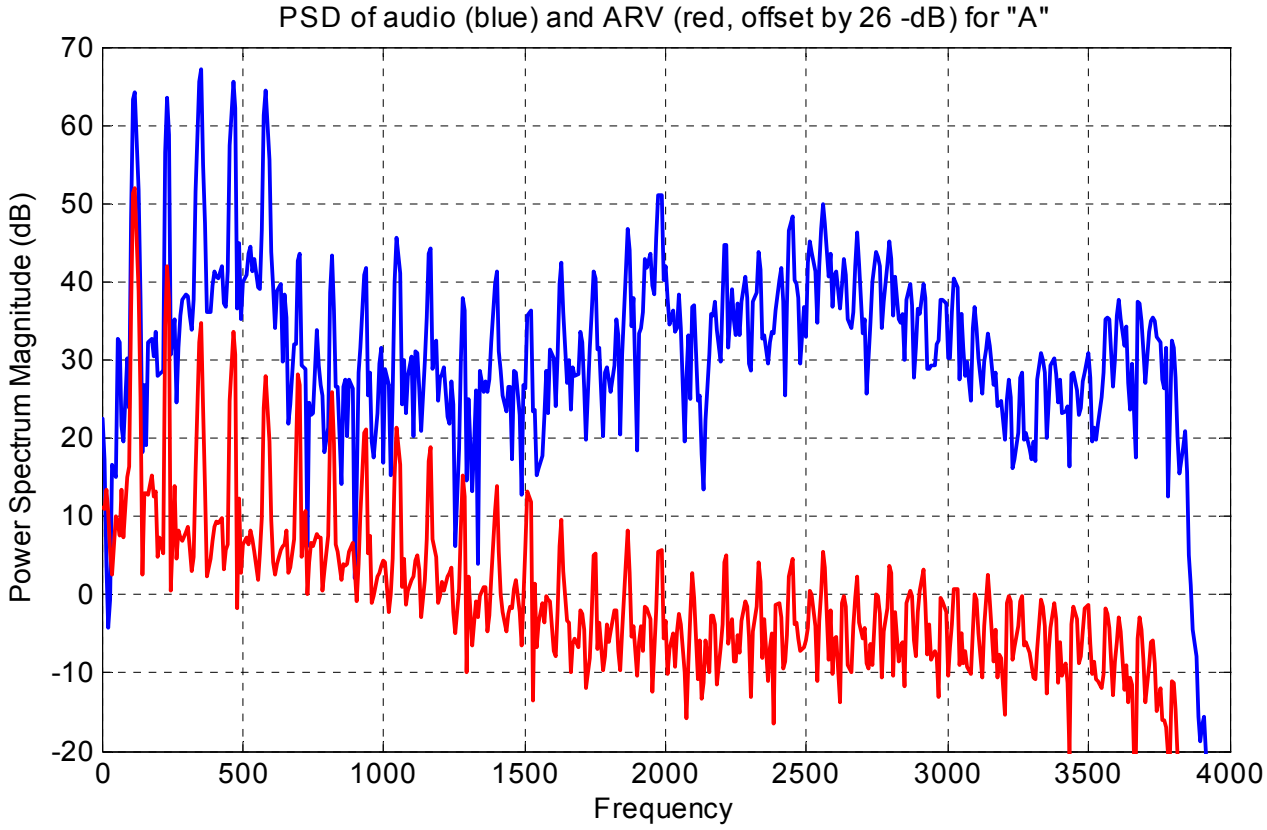

Figure 5.3. PSD of the audio (blue) and the ARV (offset by –26 dB for clarity) for the phoneme "A", or /e/. Note the lack of formant information in the ARV signal, as it is representative of the excitation and does not contain formant information. Also note the exact correlation between energy in the audio and energy in the ARV and the lack of significant ARV energy above 2 kHz.

5.3 Parametric transfer function models

The vocal tract may be modeled as a relatively complex linear time-invariant (LTI) filter that passes through certain frequencies and attenuates others. LTI systems have a linear response to an input and do not change when the input is shifted in time. These systems are well understood and form the basis for signal processing theory. Naturally LTI systems are only an approximation to the complex and nonlinear vocal tract, but the nonlinearities are assumed small and neglected for the present discussion. In addition, the vocal tract is not time-invariant, but shifts its response as different phonemes are produced. However, if the rate of change is slow, we may approximate it as a static system. This is the major impetus for processing the speech two or three glottal cycles at a time. A variety of window lengths (from 1-10 glottal cycles) have been evaluated, and a length of two cycles has been found to work well. This is often short enough so that the tract (and the pitch) changes little over this time period (10-20 msec), but long enough so that the sampled window does not contain too few samples, which causes the Fourier transforms (and the resulting models) to be too coarse. For more information on transfer function and signal processing see Section 2.4.2 of [3].

For any digital LTI system, the output y[n] can be represented as a linear combination of past inputs x[n-k] and past outputs y[n-k]:

$$A \cdot y[n] + B \cdot y[n-1] + C \cdot y[n-2] + ... = D \cdot x[n] + E \cdot x[n-1] + ...$$

where the y[n-k] are the k^{th} past output and the x[n-k] are the k^{th} past input. Transform now to frequency (z) space:

$$Y(z) \cdot \left(A + Bz^{-1} + Cz^{-1} + \ldots\right) = X(z) \cdot \left(D + Ez^{-1} + \ldots\right)$$

and since the transfer function H(z) = Y(z) / X(z)

$$H(z) = \frac{D + Ez^{-1} + \ldots}{A + Bz^{-1} + Cz^{-2} + \ldots}.$$

The roots of the numerator are the zeros of the transfer function (or plant) **H(z)** and the roots of the denominator its poles. For a further review of poles and zeros in the z-plane, please see [3], Appendix A. Parametric models are more compact than nonparametric, but sacrifice a certain amount of flexibility. Ideally, any physical system could be realized with a finite number of poles and zeros, but in reality that is rarely the case as nonlinearities make some physical systems difficult to model.

5.3.1 ARMA models

AutoRegressive Moving-Average (ARMA) models are those that incorporate both poles and zeros into the modeling process. This allows the modeling of both resonances and nulls of a system, but both the input and output of the system must be known. This is a superior model of the vocal tract because the vocal tract is not a simple tube with no nulls, but has side branches that cause zeros in the transfer function of the tract. Remember that poles are caused by the resonance frequencies of the various tubes that make up the tract, and zeros by "dead-end" tubes – those that are a branch of the vocal tract, but are closed on the other end. These "dead-end" tubes also resonate at certain frequencies, but as they are closed on the far end, the sound has nowhere to go and is eventually absorbed by the tissue. This has the effect of selectively removing certain frequencies from the vocal tract, creating minima in the output spectrum. Most of the "zero tubes" are sinuses, both in the pharynx (the piriform sinuses) and in the nasal cavities.

An ARMA model is defined by the number of poles and zeros in the system along with any delays from the input to the output (in our case caused by the travel time of the audio wave from the glottis to the microphone). The coefficients and locations of each pole and zero are then calculated so the input will match the output when operated on by the plant (to within a set tolerance).

The principal difficulty inherent in parametric modeling is that the numbers of poles, zeros, and delays of the model must be specified before processing takes place, and the number is not easily changed. As the vocal tract configuration changes for each new phoneme, the number of poles and zeros (while a good fit for some phonemes) may become inappropriate for others. Therefore it is good practice to use a few more poles and zeros than the physical situation might warrant to be sure to cover as many different vocal tract configurations as possible. A few extra poles and zeros will tend to cancel each other out (if there are enough of each), but this is known as "over-specifying" the system and can distort the transfer function with nonphysical peaks and valleys. The number of poles and zeros we use to model the vocal tract is continuously being refined, but at present is 12 poles and 6 zeros. That gives an adequate representation of most phonemes and is relatively stable over a wide variety of utterances and speakers.

5.3.2 LPC models

Linear Predictive Coding is an all-pole model. It uses past values of the output to try to predict current values. It is useful where little or nothing is known about the input to the plant. In the equation above, the numerator would consist of a single coefficient (the gain). This is because there are no known past inputs, so all the coefficients on the right side (except for D, the DC gain) would be zero. LPC is specified by the order (number of poles) and the prefilter used (the speech output is usually differentiated to emphasize the higher frequencies). LPC has been used for many years in the speech industry due to the lack of an input (i.e. an excitation function) to the vocal tract. It depends completely upon the audio (the output) to build an estimate of the transfer function of the system. As such, it is simply a compact parametric estimate of the spectrum of the output. As such, if there is little speech energy at a particular frequency with the LPC model it is not known whether that is due to lack of energy in the excitation function or a zero of the transfer function. An LPC model can locate the poles of the system

reasonably well but does not model zeros well. It is one reason LPC-based speech synthesis sounds like a person with a head cold – the zeros of the sinuses are not well represented.

5.4 Calculating the transfer function

In order to calculate accurate transfer functions using the tracheal ARV signal with fold closing feature, it is recommended to follow these steps:

1. Locate the closing feature of the ARV.
2. Segment the ARV and acoustic data into two- or three-cycle windows.
3. Use a Hamming or similar window to reduce artifacts caused by abrupt changes at the edge of the window.
4. Set the number of poles and zeros for your system to use.
5. Use a standard system identification tool (such as ARX.M in Matlab) to calculate the transfer function for that window.

Figure 5.4 (at the end of this Section) contains several calculated transfer function calculations using the ARV, pre-emphasized ARV, and pulsed VEF as excitation functions as well as an LPC model using pre-emphasized acoustic data for the phoneme “a” (/e/). All transfer functions were normalized to facilitate comparison. All models used 12 poles, and all but the LPC used 6 zeros. It is clear that the models agree quite well on formant (peak) location, but the LPC differs somewhat on the location and strength of the nulls (valleys), especially for the null near 2800 Hz. It is surprising, given the paucity of energy for the ARV above 3 kHz, that its transfer function (blue) matches the others so well. Clearly there is enough energy above 3 kHz in the ARV signal to enable the calculation of an accurate transfer function.

There is some disagreement within the non-LPC methods as to location of the zeros (especially for the null between 1000 and 1500 Hz), but overall the agreement is quite good, demonstrating that very different excitation functions

(such as the pulsed excitation and the unmodified ARV signal) can yield similar transfer functions because they have similar frequency content and are time-aligned with the acoustic data.

5.5 Combining the I and Q channels for better SNRs

Normally a simple energy calculation is sufficient to determine which channel (the **I** or **Q**) to use as the best estimate to the movement of the target, but sometimes greater SNRs are desired or both signals are relatively poor. In the latter cases, the channels may be combined[2] to yield a better overall SNR and can help minimize noise in the final signal.

Any noise present in either channel is often broadband, with similar amplitude, so we assume the noise in each channel to be of equal variance and temporally white:

$$V(t) = \begin{bmatrix} V_I(t) \\ V_Q(t) \end{bmatrix} = \begin{bmatrix} A_I \\ A_Q \end{bmatrix} \tilde{r}(t) + \begin{bmatrix} \eta_I(t) \\ \eta_{Q(t)} \end{bmatrix}$$

where $\tilde{\mathrm{r}}(\mathrm{t})$ is the signal of interest and $\eta_{\mathrm{I}}(\mathrm{t})$ is the noise in the **I** channel. We now seek to obtain the "best" linear combination of the two channels. We define "best" as the following:

1. The linear combination of interest is $o(t; w_I, w_Q) = w_I V_I(t) + w_Q V_Q(t)$, and the objective is to find the best values for w_I and w_Q under the constraint $w_I^2 + w_Q^2 = 1$.
2. Define the estimated covariance matrix for the process as:

$$\hat{R} = \hat{R}_d + \hat{R}_n = \frac{1}{T}\int_0^T dt V(t) V(t)^T$$

3. Suppose that the noise is equal-variance and uncorrelated, so that

[2] Again, thanks to Dr. Samuel L. Earp of Multisensor Science, LLC, for assistance in this section.

$$\hat{R}_n \approx \sigma^2 I$$

$$\hat{R}_d \approx \left(\frac{1}{T} \int_0^T dt \tilde{r}^2(t) \right) \begin{bmatrix} A_I^2 & A_I A_Q \\ A_I A_Q & A_Q^2 \end{bmatrix}$$

Then

4. The maximum output SNR is for a weight vector

$$[w_I \quad w_Q] = \begin{bmatrix} A_I & A_Q \end{bmatrix} / \sqrt{A_I^2 + A_Q^2}$$

5. This weight vector is approximately the eigenvector associated with the largest eigenvalue of $\hat{R}$, which is easily computed.

Hence the "best linear combination", under these assumptions, is easily computed, and the best output SNR is given as

$$SNR_{OPT} = \frac{(A_I^2 + A_Q^2) E_{\tilde{r}}}{\sigma^2}$$

$$E_{\tilde{r}} = \frac{1}{T} \int_0^T dt \left[\tilde{r}(t) \right]^2$$

Notice that if the coefficients are such that there is significantly more energy in one channel than the other, then there is little loss in just selecting the channel with the most energy. The result of this computation should be equivalent to automatically selecting the channel with the most energy. For most situations, this is probably not worth the computational effort, but for others, it may well be.

To actually increase the SNR of the system, the noise of the system must be addressed. A standard one-channel spectral subtraction system is useful for this, as the noise is normally much lower in energy than the signal of interest, but a pre-whitening technique may also be employed. Instead of using the eigenvalues of **R**, as above, we use the eigenvalues of $[\mathrm{R}_\mathrm{n}^{-1/2} \mathrm{R} \mathrm{R}_\mathrm{n}^{-1/2}]$, which results in the suppression of some noise from the system, but requires an

estimate of $R_n^{-1/2}$. This is not overly cumbersome, but a spectral subtraction technique is probably a simpler option.

5.6 Conclusion

The ARV is capable of providing accurate voiced speech detection, pitch, and an excitation function in any acoustical environment. Its signal can be used to calculate ARMA models of the transfer function of the vocal tract, but as with conventional models (such as LPC) this requires relatively noise-free acoustic speech data. ARMA models also require specification of the number of poles and zeros of the system under consideration. Unlike LPC, though, which is simply a compact representation of the output power spectrum, the ARMA models are capable of accurately simulating the nulls of the system, resulting in more natural sounding synthesized and coded speech. An excitation function composed of pulses where the closing (and optionally the opening) of the folds occurs is probably more robust than using the ARV signal alone, although either can be used and the ARV signal requires less preprocessing.

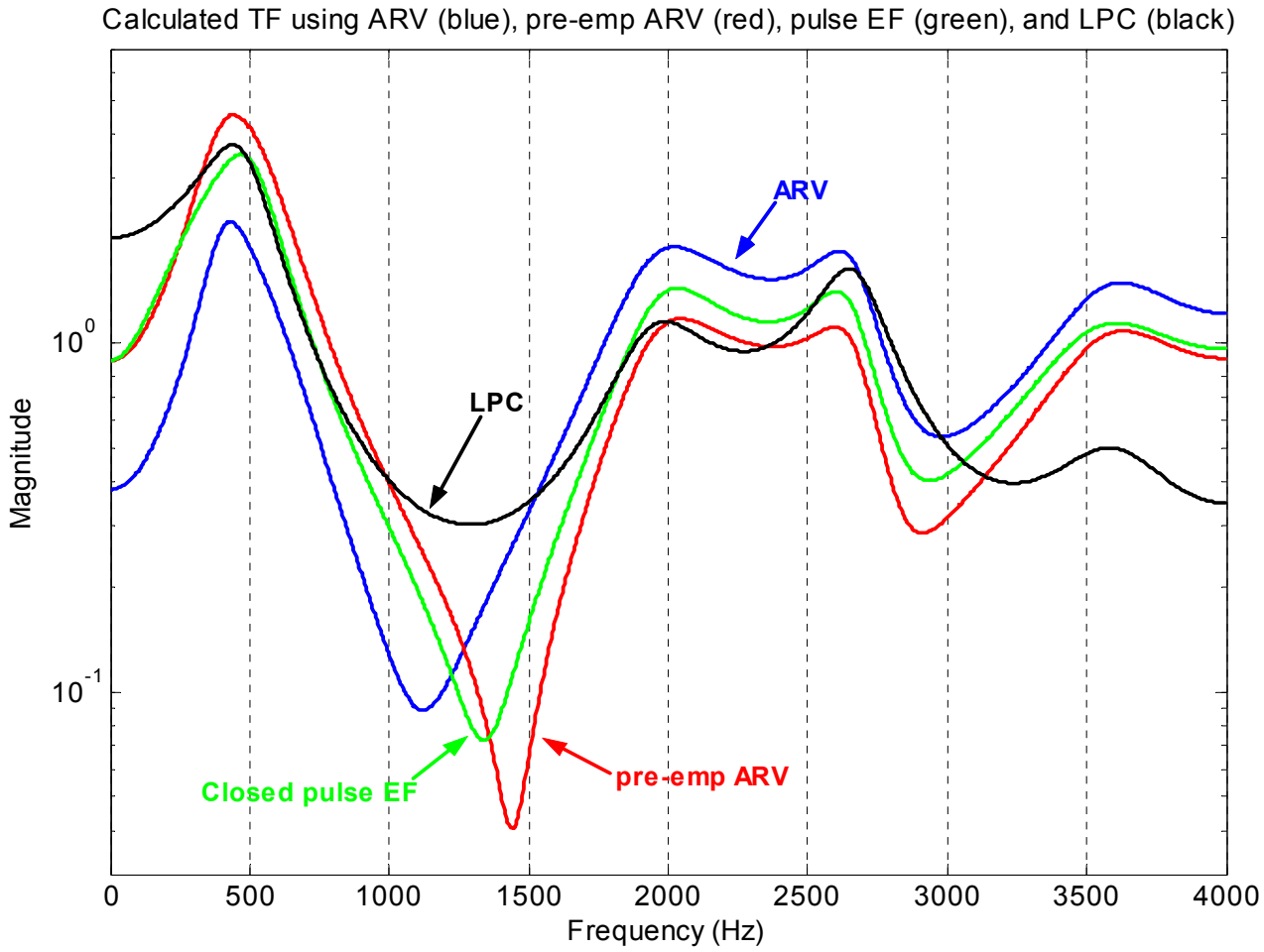

Figure 5.4. Several different calculated transfer functions for the phoneme "A" (/e/) using the ARV (blue), the pre-emphasized ARV (red), the closed-pulse excitation function (green), and a pre-emp LPC model. All models used 12 poles and all but the LPC used 6 zeros. This was calculated over a 2 glottal-cycle window (about 128 samples at 8 kHz). Note that all techniques basically agree on formant locations (peaks) but that the LPC does not model the zeros (valleys) as effectively as the other models, especially the one near 2800 Hz.

6. Limitations of the ARV

The ARV is a useful tool that can be used to generate data not previously available (including very high noise environments), but does have limitations:

What you can use the ARV to do:

1. Determine when the user is speaking, even in heavy noise
2. Determine the pitch of the user, very accurately and every single glottal cycle if desired, even in heavy noise
3. Construct an accurate estimate of a voiced excitation function, even in heavy noise
4. Construct an accurate model of the vocal tract, including zeros

What it is difficult or impossible to do with the ARV:

1. "Image" the trachea or vocal folds – the ARV is simply a vibration detector
2. Determine the absolute amplitude of motion of the target
3. Determine EXACTLY what is causing the signal in the presence of many targets – the ARV detects and superposes all movement within its "field of view", and this can include both near and far field interactions – although normally not at the same time.

Physical or practical limitations

1. The cable is somewhat fragile and when frayed can cause ARV noise
2. The antenna has to be matched to the skin of the user, carefully placed on the skin, and remain in the same relative position throughout a measurement
3. Environmental vibration can affect the ARV return and lead to broadband ARV noise

The ARV user should not look upon these limitations as an impediment to producing good data with the ARV, but as a guideline to efficient ARV use.

7. Conclusion

The Aliph RadioVibrometer is a homodyne interferometer that detects vibratory motion of targets in a variety of environments, including inside the human body. As long as certain restrictions regarding relative amplitude of target motion are followed and both output channels examined, an accurate representation of the motion of the target is possible. The ARV detects motion using an antenna, which may be a conventional free-space antenna if detection is desired at a distance in air, or a high-dielectric antenna designed to operate well on the human body. When placed on the skin, the far-field operation of the antenna is disrupted and the primary interaction with the tissue is through the near field. The ARV can detect vibratory motion anywhere vibrating tissue is located close enough to the surface to interact with the near field of the antenna. Operation of the ARV near the laryngeal prominence results in a superposed signal containing both vibratory and vocal-fold closure elements of roughly the same amplitude. For speech applications, the ARV signal can be used in any acoustic noise environment to determine the presence of voiced speech, very accurate pitch, and a voiced excitation function. With the recorded speech, it may also be used to calculate an accurate transfer function model of the vocal tract, which can be used in many applications, including speech synthesis and coding. The two channels of the ARV may be combined or the largest energy channel signal chosen for use in the above applications.

8. References

[1] *The Electronics Handbook* (1996), pp 1191-1193, Jerry C. Whitaker, ed. CRC Press. ISBN 0-8493-8345-5.

[2] http://www.sospubs.co.uk/sos/may00/articles/synth.htm

[3] "The Physiological Basis of Glottal Electromagnetic Micropower Sensors (GEMS) and their Use in Defining an Excitation Function for the Human Vocal Tract", Gregory C. Burnett, Ph.D. Thesis, U.C. Davis, 1999 (paper copies available at UMI: http://www.umi.com/hp/Products/Dissertations.html, thesis # 9925723; or online at http://www.aliph.com/burnett_thesis/).

[4] Duck, F.A. (1990). "Physical Properties of Tissue" (Academic Press Inc., San Diego, CA), Table 6.13.

[5] Lin, JC (1986). "Microwave propagation in biological dielectrics with application to cardiopulmonary interrogation", *Medical applications of microwave imaging*. New York: IEEE Press, ISBN: 0-87942-196-7, pp.47-58.

[6] Griffiths, David J. (1989). "Introduction to Electrodynamics" (Prentice-Hall, Englewood Cliffs, NJ) 2nd edition, p. 271.

[7] Haddad, W.S., Rosenbury, E.T., Johnson, K.B., and Pearce, F.J. (1997). Measurements of the Dielectric Properties of Body Tissues and Fluids at Microwave Frequencies (Lawrence Livermore National Laboratory, to be published)

[8] Titze, Ingo R. (1994). "Principles of Voice Production" (Prentice-Hall, Englewood Cliffs, NJ)

[9] http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/voluntary.html