Version 1.0 – 14 November 2004

1 The role of the laryngeal collar in vocal tract acoustics

Wim G. J. Ritzerfeld, M.Sc., Donald G. Miller, M.Mus, Ph.D,
and Harm K. Schutte, M.D., Ph.D.

Groningen Voice Research Lab. University of Groningen, Ant. Deusinglaan 1,
NL 9713 AV, Groningen, The Netherlands

1.1 Abstract

In "Acoustic Interpretation of Resonant Voice", Titze has presented a theory stating that the vibrational movements of the vocal folds can be reinforced by vocal tract resonance (2001). In order for this reinforcement to take place, vocal tract impedance needs to have a positive phase angle. In other words, the supraglottic pressure wave needs to be pulled forward in time with respect to the glottal airflow wave. Titze presented calculations indicating that this favorable condition is enhanced when: the passage to the pharynx from the vestibule via the laryngeal collar or epilaryngeal tube (ET) is narrow.
In the present study, experiments were carried out that were aimed at investigating the above theory empirically. A subject (DM) produced several phonations, which were simultaneously measured by a pressure transducer located inside the ET, just above the vocal folds, an electroglottograph, and a microphone placed in front of the subject. Results were established that support the conclusion that just above the vocal folds the acoustical wave shows the theoretically desired phase shift. The experiment provided support for the theory that there is a correlation between this phase shift and the sound pressure level. However, The question whether or not feedback to the source is responsible for the measured difference in SPL could not be answered.

1.2 Introduction

The *laryngeal collar* or *epilaryngeal tube* (ET) plays an important role in vocal tract acoustics, especially in the generation of the singer's formant (SF). In 1974 Sundberg [1] wrote that a clustering of the higher formants into the singer's formant takes place if the cross-sectional area of the ET is at least a factor of six smaller than the cross-sectional area of the pharyngeal space directly above it.

A less well known effect of a narrowed ET is that the *impedance* (acoustic load) of the vocal tract is modified in such a way that a change occurs in the time relation between the air puffs through the glottis and the sound wave above the glottis, whereby the maxima and minima of the sound wave appear *earlier* in time. Such type of impedance is referred to by electrical engineers as *inductive* and by mechanical engineers as *inertive*.

Several authors have suggested that this *phase shift* of the supraglottic sound wave promotes positive feedback to the *voice source*.
Rothenberg [1] showed that the glottal airflow pulse (the air puff) can be measured indirectly by using an airflow mask, put over the mouth of the subject, to measure oral airflow and, by inverse filtering, derive the glottal airflow signal. Results indicate that the airflow pulse is skewed and that this skewing cannot be explained by the change of glottal area during the open phase of the glottis. For chest voice he found that the airflow pulse is skewed to the right, with the maximum appearing shortly before the moment of closing of the vocal folds.

He was able to show theoretically that this skew to the right of the glottal airflow pulse could result if the abovementioned phase shift of the supraglottic sound wave occurs.
Titze [3] suggested that the phase shift could have a positive effect on *vocal fold oscillation* because it increases the driving force on the vocal folds during opening as well as during closing.

In this paper we will first present the theory of vocal tract impedance. We will then present the results of a series of measurements of *supraglottic pressure.*

1.3 The inductive vocal tract

Rothenberg [1] explained the skew to the right of the glottal airflow pulse by assuming vocal tract impedance to be inductive. Inductive vocal tract impedance by definition causes supraglottic pressure to be proportional to the rate of change of glottal airflow. Note that an electric coil behaves in exactly the same manner. This condition can be expressed as follows:

(Eq.1) $$P_{supra} = L\frac{dU}{dt}$$

where:
P_{supra} = supraglottic pressure
L = (acoustic) inductance
U = volume flow

The effect of this type of impedance on the supraglottic sound wave is that the maxima and minima of this sound (pressure) wave appear earlier in time than in the situation where supraglottic pressure is simply proportional to air flow itself. The effect on the glottal airflow pulse is that, as the vocal folds separate, the rise of airflow is being slowed down, since a higher rate of change of airflow will result in a higher supraglottic pressure, which will in turn limit the airflow. The maximum of the airflow pulse will thus appear late in the cycle, shortly before the moment of closing, where airflow will be cut off suddenly by the closing vocal folds. The rate of change of airflow (and thus Psupra) will then first become strongly negative and, shortly after that go to zero. We thus expect a sharp negative impulse of Psupra at the closing moment (indicated by the vertical arrow in Figure 1). The vocal tract is not purely inductive, however. Apart from its inductive behavior, it also has a number of discrete resonances, called formants. The vocal tract will thus react to the aforementioned impulse by its *impulse response*: a superposition of (damped) sine waves at the formant frequencies. Note that the fact that the glottis is now closed makes it very effective acoustically, also at high frequencies (*singer's formant*). Rothenberg has compared this mechanism of the slowly rising airflow, which is suddenly cut off, to the operation of the spark coil in an automobile, where the electric current is first allowed to rise to a high value and then is suddenly cut off, producing a high voltage pulse, which is used to produce the spark for igniting the fuel.

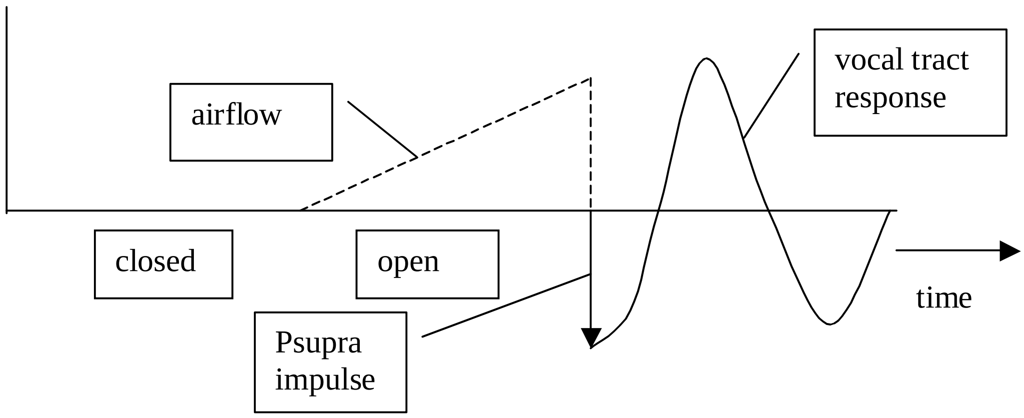

Figure 1 Glottal airflow pulse and resulting P_{supra} impulse

1.4 Calculation of vocal tract impedance

1.4.1 Input impedance of a cylindrical closed-open pipe

A cylindrical pipe, which is closed at its input and open at its output, exhibits resonance peaks (formants), due to reflections at the closed and open ends. Due to these reflections, it will also have a frequency-dependent input impedance, which will contain peaks at these formant frequencies.

According to Fletcher & Rossing [5], the input impedance of an open-ended, loss-less cylindrical pipe can be calculated using the following equation (8.23):

(Eq.2) $$Z_{in} = \left[\frac{Z_L \cos kL + jZ_0 \sin kL}{jZ_L \sin kL + Z_0 \cos kL} \right]$$

where:

Z_{in} = input impedance
Z_L = load impedance (zero for an ideal open end, infinite for a stopped pipe)
Z_0 = Characteristic impedance; $Z_0 = \rho c/S$, where ρ=density of air, c=speed of sound, S is cross-sectional area
k = wave number = $2\pi/\lambda = 2\pi f/c$ (λ=wavelength, f=frequency, c=speed of sound)
L = length of the pipe

Realistic pipes will not be loss-less. They will have essentially three types of losses:

1. *Radiation losses*. Sound will emit from the open end and the acoustic power emitted will be lost to the internal resonance of the pipe. Therefore, the load impedance at the open end will not be zero (see [5], section 8.3).
2. *Wall losses*, due to viscous drag and thermal exchange between the air and the walls (see [5], section 8.2).
3. Mechanical *vibrations of the walls*.

1.4.2 Calculation

(Eq.2), along with the equations for radiation losses and wall losses in [5], sections 8.2 and 8.3 were entered into a spreadsheet. The third factor, mechanical losses of the walls, was neglected. The so-called low-loss approximation was used, meaning that Z_0 is still a *real* number, but the wave number *k* becomes a *complex* number.

Note that correct incorporation of wall losses is not trivial and the calculations presented here must be seen as approximations. In order to get more accurate results, the exact nature of these losses and the proper values of the parameters involved would have to be studied in more detail.

The result can be seen in Figure 2. Impedance (magnitude) is plotted in kg/m^4s, phase in degrees and frequency in Hz. The length of the pipe is 17,2 cm and its cross-section is 6 cm^2.

Some conclusions can be drawn from Figure 2:

- The formant frequencies are approximately 466 Hz, 1398 Hz, 2330 Hz etc.. Note that the formant frequencies are lower than would be expected from the length of this tube. With c=344 m/s, F1 would be 344/(4*0,172)=500Hz instead of 466 Hz. The difference is due to the radiation losses, which make the pipe's acoustical length greater than its physical length by approximately 0,85.r where r is the radius of the pipe (see [5], eq. 8.32).
- The magnitude of the formants decreases with frequency.
- The phase of Z is zero at the formant frequencies.
- The phase of Z is positive below the formant frequencies and negative above the formant frequencies.

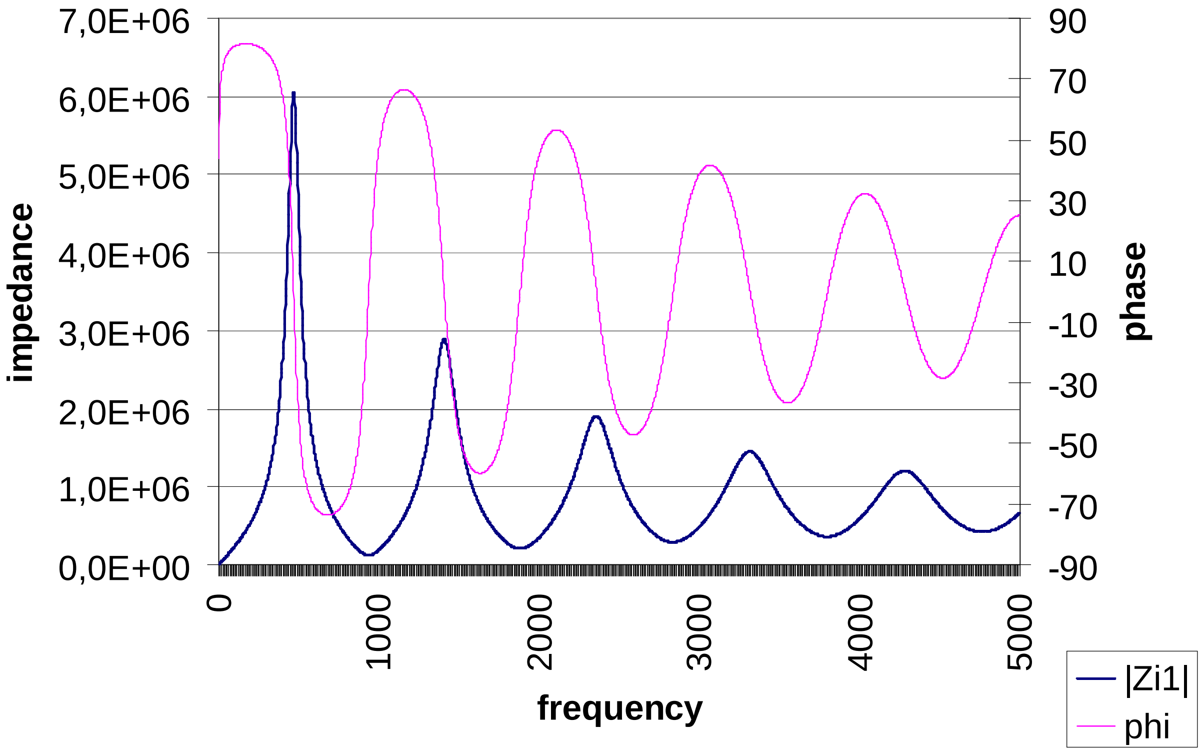

Figure 2 Impedance (magnitude and phase) of a lossy cylindrical pipe with losses

1.5 Input impedance with narrowed epilaryngeal tube

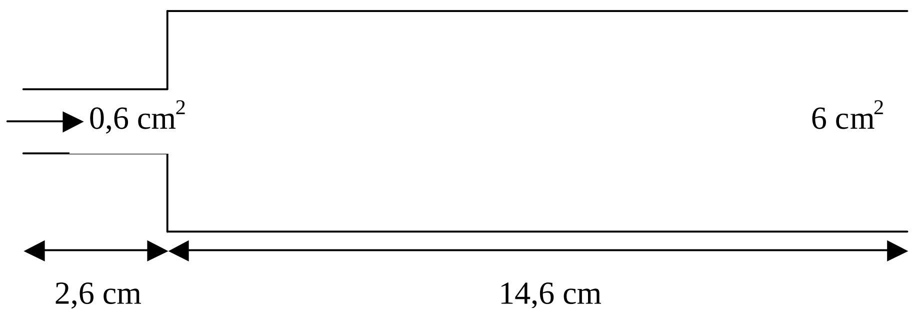

Figure 3 Cylindrical pipe with narrowed epilaryngeal tube

In order to calculate the input impedance of the pipe shown in Figure 3, we have to work in two steps. For every frequency we first calculate the input impedance of the wide section of the pipe, using (Eq.2). Then we fill in this impedance as load impedance (Z_L) into (Eq.2) for the narrow section of the pipe. This will give us the total input impedance for that frequency.

The result of the calculation is plotted in Figure 4.

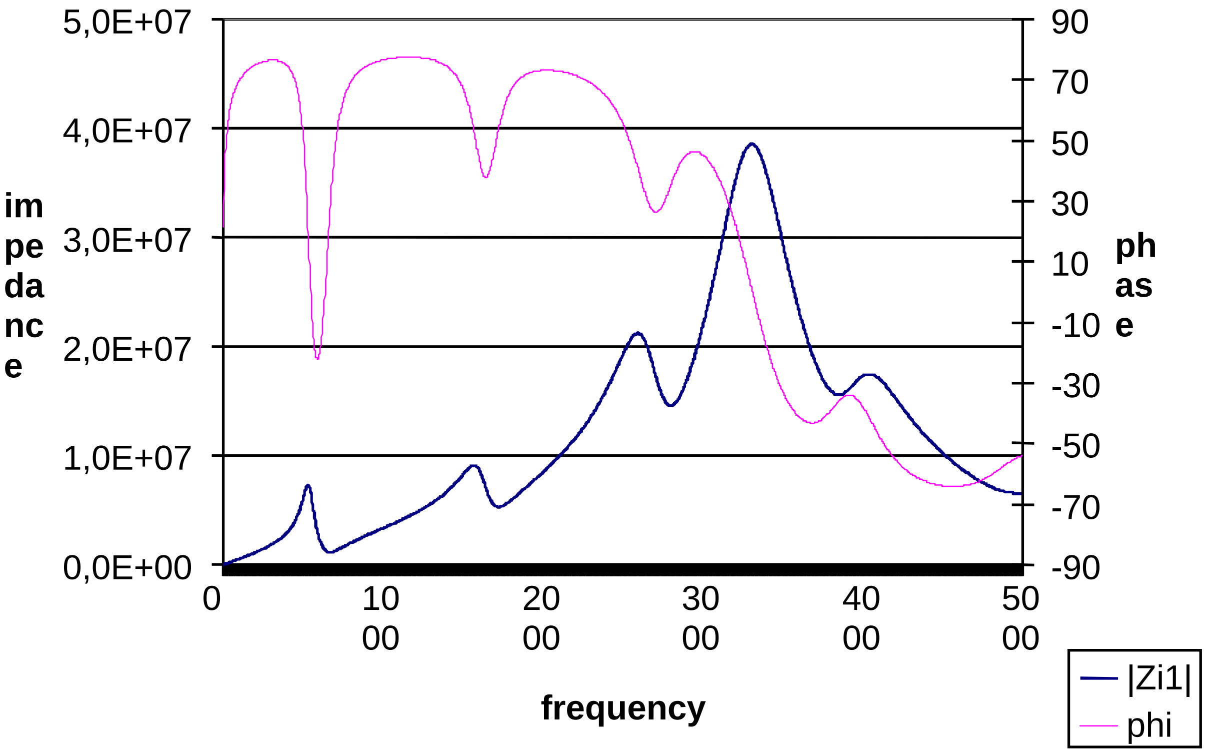

Figure 4 Impedance (magnitude and phase) of a lossy cylindrical pipe with narrowed epilaryngeal tube

The following conclusions can be drawn:

- The magnitude of the formants increases with frequency up to about 3500 Hz and then decreases.
- F1 and F2 have gone up to 528 Hz and 1566 Hz respectively. Apparently, the acoustic length of the pipe decreases due to the narrowed epilaryngeal tube.
- The phase of Z is positive for most frequencies, except for a small range above F1. Even exactly at the formant frequencies, the phase of Z is still positive: 24 degrees at F1 and 54 degrees at F2).

What we are seeing here is that the pipe has become *inductive*, but it does not quite conform to (Eq.1). If pressure would be proportional to the rate of change of airflow, the phase of Z would have to be 90 degrees, independent of frequency and the magnitude of Z would have to rise linearly with frequency. This can be verified as follows:

Suppose $U(t) = A.\sin(\omega t)$, then according to (Eq.1), P(t) would have to be:

$$P(t) = L\frac{dU(t)}{dt} = L.A.\omega.\cos\left(\omega t\right) = L.A.\omega.\sin\left(\omega t + 90^o\right)$$

meaning that the pressure amplitude rises linearly with frequency and the phase is $+90^o$ for all frequencies. What we see in Figure 4 is just an approximation of this ideal. We can think of our cylindrical pipe with narrowed ET as a resonator that is inductive, but on top of that inductance it still has formants, although these are less prominent than they are without the narrowed ET.

Although the pipe has not become purely inductive, we now know that the phase of Z is positive for most frequencies. This means that the pressure maxima and minima appear earlier in time than the airflow maxima and minima for most frequencies, except in a small range above F1.

1.6 The time domain

In order to explain some of the subtleties of the interaction between the voice source and the vocal tract, it is necessary to look at signals in the *time domain* rather than the *frequency domain*, as we have done in the previous sections.

In order to calculate the resulting P_{supra} from the glottal pulse signal, we could derive the differential equation of the system, fill in the source signal function and solve the equation for P_{supra}. This is not a trivial task if one has to do it manually, but it could be done using e.g. an electronic circuit simulator or a mathematics package.

An alternative method would be to calculate the Fourier series of the glottal pulse signal. Multiplying the Fourier coefficients with the values of vocal tract impedance at the relevant frequencies would give us the Fourier series of P_{supra}. This is also not a particularly simple task.

Nonetheless, we can still predict what the time domain signal will look like, applying the theory derived thus far. A recipe to do this consists of the following steps:

1. Assume a certain shape for the airflow pulse.

2. Take the time derivative (rate of change) of this pulse to obtain the shape of $\mathbf{P_{supra1}}$, which is the supraglottic pressure **caused directly by the airflow pulse**, ignoring any oscillations or standing waves.
3. Now assume there is a sinusoidal oscillation $\mathbf{P_{supra2}}$, **caused by resonance in the vocal tract**, which has the same frequency as the dominant formant. In general, this oscillation is not just the result of the vocal tract reacting to the most recent closing of the vocal folds, but it also contains the energy, which has been conserved during previous cycles. The closer the formant matches the relevant harmonic, the more this will be so and the more this oscillation will have the characteristics of a *standing wave*.
4. Once this oscillation enters the open phase of the glottal cycle, it interferes with P_{supra1}. Effectively, P_{supra1} and P_{supra2} must be added together in the open phase to obtain the total P_{supra}. Be aware of the fact that damping is higher in the open phase, due to the open glottis absorbing part of the energy, especially at high frequencies. Take into account that the airflow pulse can be slightly modified by the resulting P_{supra} and start again with 1.

1.7 Measurement results

1.7.1 Description

The experimental setup calls for a Millar catheter with an outside diameter of 1.6 mm, and pressure transducers at the tip and 6 cm from the tip. The catheter is inserted through the nose and into the posterior commissure of the glottis in such a way that the pressure sensors, without mechanical contact with bodily parts, can measure the time-varying pressures in the spaces just above and below the glottis, respectively. Additional measured signals are from a microphone 30 cm in front of the subject and from an electroglottograph (Fourcin). The four signals were recorded on an instrumentation recorder (Racal) at a tape speed of 60 i.p.s.

The subject, one of the authors (DM), is a bass-baritone with 25 years of professional singing experience.

The original data is taken with calibrated pressure transducers, and from this data it was shown that mean subglottic pressure is effectively constant throughout the phonation. For the purposes of this study, the signal from the transducer above the glottis, designated Psupra, is normalized with respect to amplitude and treated as a microphone signal. Here the phase of the variations of (acoustic) pressure is of principal interest, and the software VoceVista is used for making delay corrections and display.

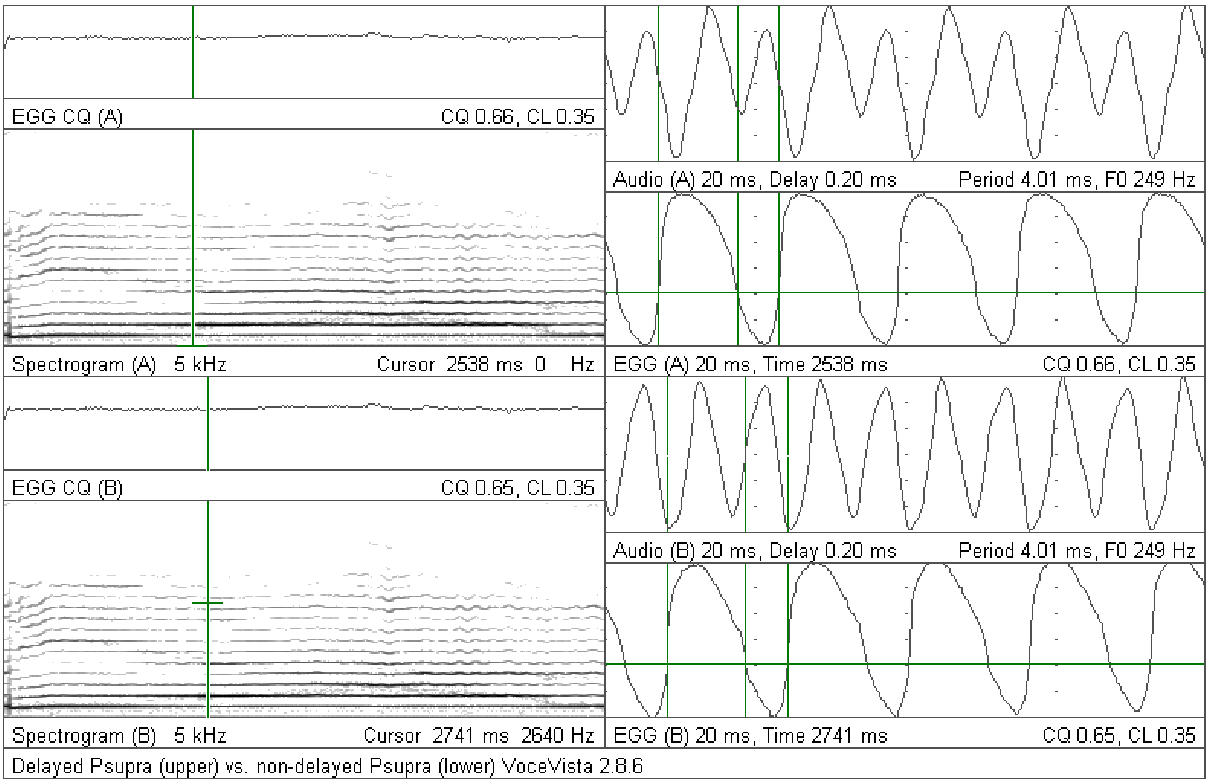

Figure 5 Vowel series: delayed Psupra versus non-delayed Psupra (vowel [e])

Figure 5 shows the difference between the supraglottic pressure wave at two points in time during the first half of the vowel series (vowel approximately [e]) sung on pitch G3. The vowel series is executed in such a way that the first formant (F1) gradually rises during the first half of the exercise and gradually falls during the second half. Both points in time occur during the first half, where F1 is rising.

1. In the upper panel (Audio window) we see that the first minimum in P_{supra} appears *after* the moment of closing of the vocal folds. The moment of closing can be identified as the moment when the electroglottograph (EGG) signal exhibits a sharp rise.
2. In the lower panel (Audio window) we see that some time later the minimum in P_{supra} coincides with the moment of closing of the vocal folds.

Note that all signals in the right half of Figure 5. are normalized to full scale by VoceVista, meaning that no conclusions can be drawn from the amplitudes of these signals.
Figure 6 shows the power spectrum and sound pressure level at the same two points in time during the vowel series where the upper and lower panels in the figure correspond to the upper and lower panels in Figure 5.
It can clearly be seen that:

1. In the upper panel the sound pressure level (SPL) is low.
2. In the lower panel the sound pressure level is high.

Note also that the greatest difference in amplitude occurs in the second (H2) and fifth (H5) harmonic.

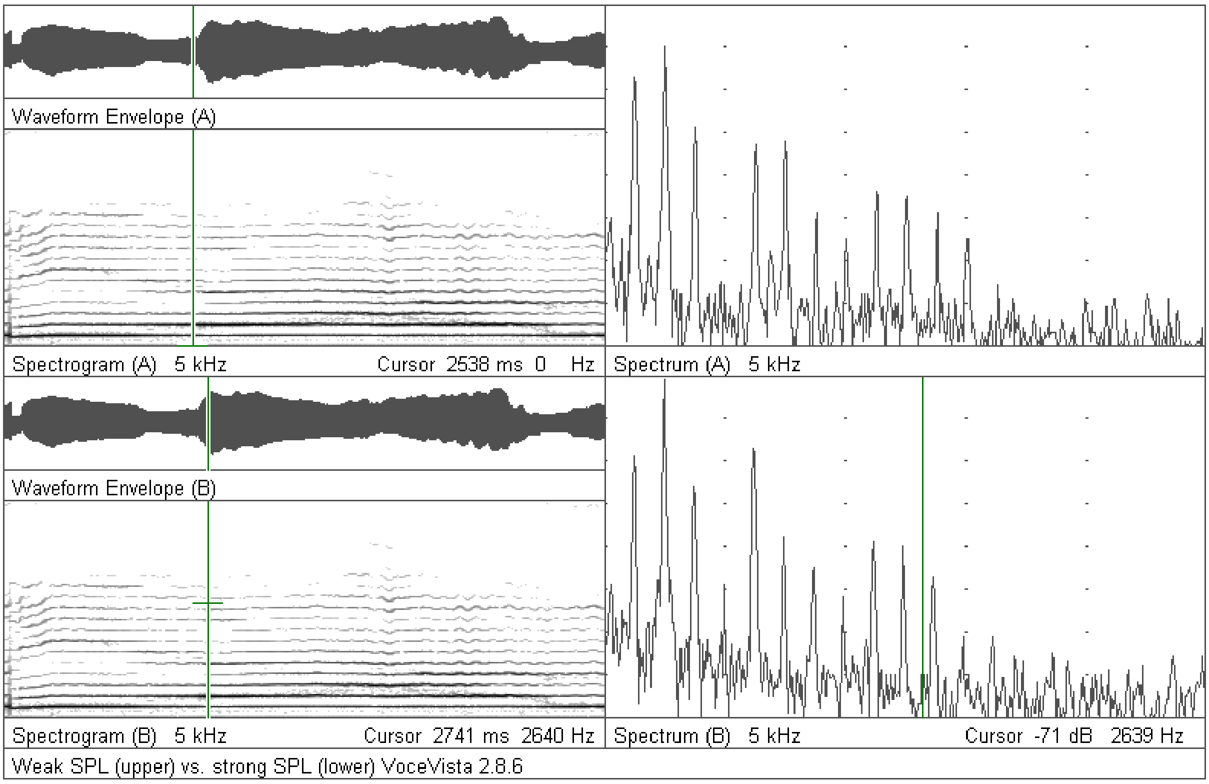

Figure 6 Vowel series: weak SPL versus strong SPL (vowel [e])

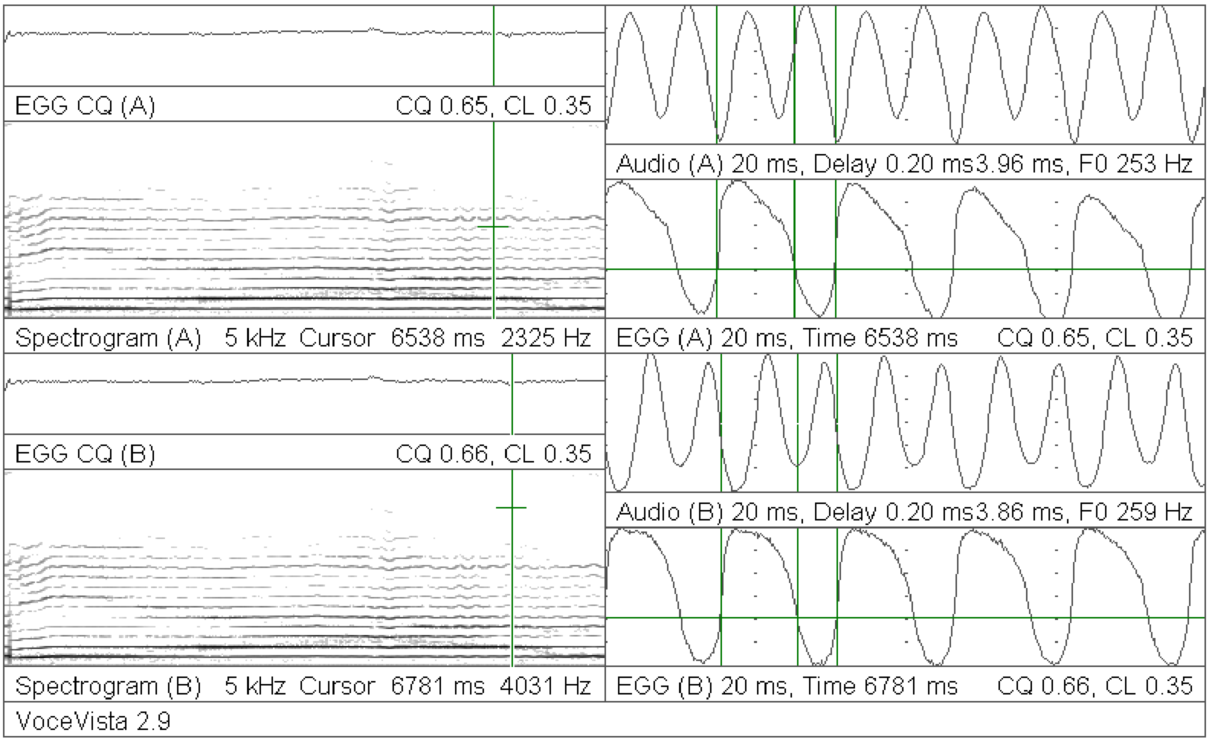

Figure 7 Vowel series: delayed P_{supra} versus non-delayed P_{supra} (vowel [o])

Figure 7 shows the difference between the supraglottic pressure wave at two points in time during the second half of the vowel series (vowel approximately [o]).

3. In the upper panel (Audio window) we see that the first minimum in P_{supra} appears at the moment of closing of the vocal folds.

4. In the lower panel (Audio window) we see that some time later the minimum in P_{supra} appears *after* the moment of closing of the vocal folds.

Figure 8 shows the power spectrum and sound pressure level at the same two points in time during the vowel series where the upper and lower panels in the figure correspond to the upper and lower panels in Figure 7.
It can clearly be seen that:

3. In the upper panel the sound pressure level (SPL) is low.
4. In the lower panel the sound pressure level is high.

Again the greatest difference in amplitude occurs in the second (H2) and fifth (H5) harmonic.

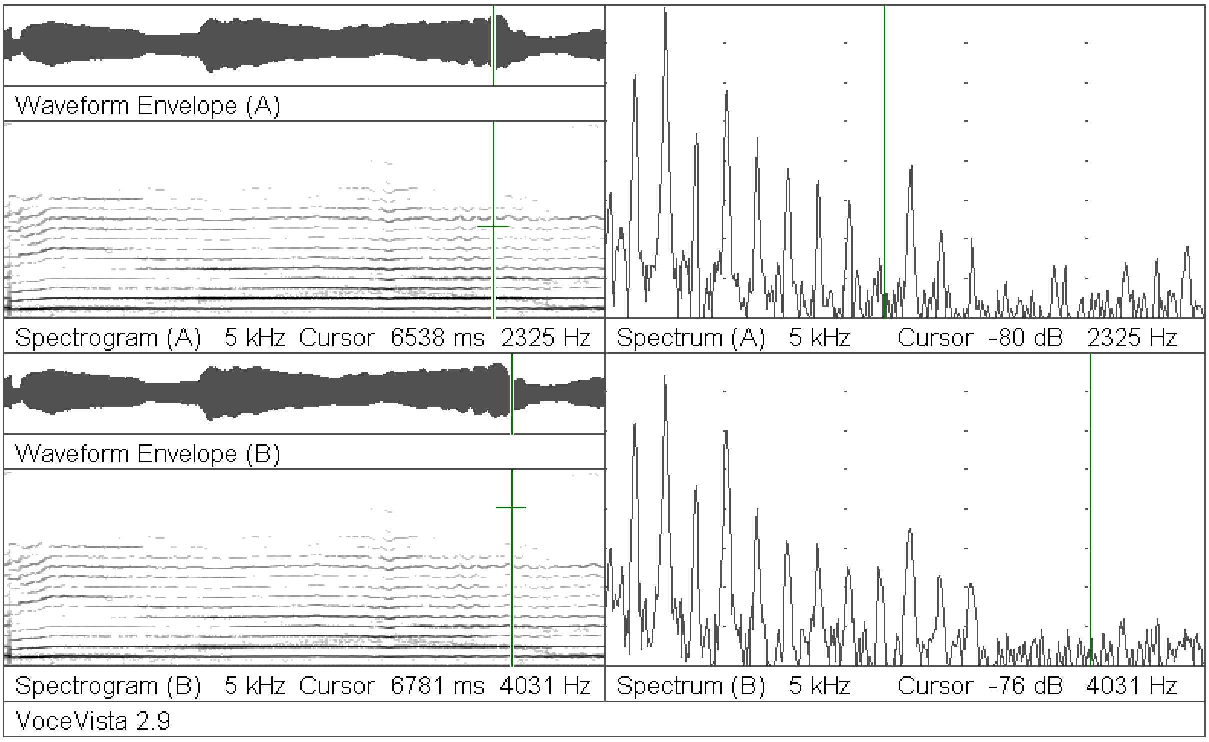

Figure 8 Vowel series: weak SPL versus strong SPL (vowel [o])

1.7.2 Discussion

What we see happening here is a phase shift that is associated with the tuning of the first formant (F1) with respect to the second harmonic (H2). As we have seen in Figure 3, the phase of vocal tract impedance is negative within a small frequency region above F1. This means that whenever a harmonic of P_{supra} has a frequency that lies within this region, it is delayed with respect to the corresponding harmonic of the airflow signal. This is what causes the second maximum in P_{supra} to appear late in the cycle and the subsequent minimum to shift into the following cycle.
It looks as if this delay is responsible for the much lower sound pressure level. The question now is: what is the underlying physical cause of this difference in SPL?
We see three possible models:

- A high P_{supra} near the end of the cycle and a low P_{supra} during opening of the glottis will destroy the skewing of the glottal airflow pulse (the maximum will not appear at the end of the cycle anymore) and thus the rate of change of airflow at the end of the cycle will become less negative. This diminishes the “spark plug” effect as described by *Rothenberg* (see 1.3) and thus the oscillation after closing will be weaker. According to Rothenberg, a less skewed airflow pulse can also be expected to contain less high harmonics than a skewed pulse. In Figure 6 and Figure 8 there is a difference of a few dB in amplitude of the harmonics around 2500 Hz, but most of the difference can be found in the lower harmonics.

- According to *Titze*, the high P_{supra} near the end of the cycle will increase *intraglottal pressure* (the pressure within the glottis on the vocal folds) and this will slow down the closing of the vocal folds. Similarly, the low pressure during opening of the glottis will slow down the opening of the vocal folds. Thus vocal fold oscillation is damped and the result is a lower SPL. If vocal fold oscillation is damped, one would also expect the amplitude of the higher harmonics to become weaker, but, as stated already, the difference in amplitude of the higher harmonics in our data is much smaller than for the lower harmonics (see Figure 6 and Figure 8).
- The third model does not assume any interaction between the acoustic pressure in the vocal tract and the voice source. It can be described using the method outlined in 1.6: The pressure component that is directly generated by the airflow pulse (and which we called P_{supra1} in 1.6) generally has a minimum just before closing and a maximum shortly after opening. This can easily be verified by examining the rate of change of airflow. The rate of change of airflow is highest shortly after opening, resulting in a maximum, and lowest (most negative) just before closing, resulting in a minimum. Now the resonant part of P_{supra} (which we called P_{supra2} in 1.6) must be added to P_{supra1} to get the total P_{supra} in the open phase. In order for the amplitude of P_{supra} to be maximal, the maxima and minima of P_{supra2} must coincide as much as possible with those of P_{supra1}. This is the case for both the maximum and the minimum in the "strong" portion of the example, but not in the "weak" portion and thus the SPL in the weak portion is diminished.
  Note that this model would explain why most of the difference in SPL is caused by the difference in amplitude of the second harmonic.

1.7.2.1 Conclusion

We have been able to show empirically that the phase of P_{supra} is dependent on formant tuning.

We have not been able to conclude which of the above models adequately explains the large difference in SPL. Therefore, the question whether or not feedback to the source is responsible for this difference must remain inconclusive.

1.8 References

[1] Sundberg, J. Articulatory interpretation of the "singing formant", in Journal of the Acoustical Society of America Vol. 55, p. 838-844, 1974

[2] Rothenberg, M. "Acoustic Interaction Between the Glottal Source and the Vocal tract", in Vocal Fold Physiology, K, N. Stevens and M. Hirano, Eds., University of Tokyo Press, pp. 305-328, 1980.

[3] Titze, I.R. "Acoustic Interpretation of Resonant Voice", in Journal of Voice, Vol. 15 no. 4, pp. 519-528, 2001.

[4] Titze, I.R. "Principles of Voice Production", Prentice-Hall 1994

[5] Fletcher, N.H., Rossing, T.D. "The Physics of Musical Instruments"