EXPLORING THE IMPORTANCE OF FORMANT BANDWIDTHS IN THE PRODUCTION OF THE SINGER'S FORMANT

Tom Millhouse[1], Frantz Clermont[2] and Pamela Davis[1]

[1] National Voice Centre, University of Sydney
[2] University College, (ADFA Campus), University of New South Wales

ABSTRACT-This paper presents some preliminary work in observing the formant frequency bandwidths of the singing voice. It discusses mathematically how the reduction of specific formant bandwidths can increase the level of the singer's formant and presents some experimental data observing narrow bandwidths as a possible factor in the production of the singer's formant in a single classically trained bass-baritone subject. Bandwidth information was not calculated directly but inferred from the glottal reflection coefficient and a selective glottal reflection coefficient.

INTRODUCTION

The singer's formant is a special resonatory phenomenon that occurs in western, classically trained opera singers (Sundberg, 1974). It makes the voice more audible to the human ear even in the presence of a loud orchestral accompaniment. Opera singers have been shown to increase the power amplitude of their voice to the more frequency sensitive regions of the aural spectrum (Sundberg 1974), thus allowing the listener to distinguish more easily between the orchestra and the singer. The singer's formant has been observed in all lower pitched vocal types: bass, baritone, tenor and alto (Bloomooft & Plomp, 1986). Acoustically, this resonatory phenomenon has been observed has a clustering of the third, fourth and fifth formant resonant frequencies (F_3, F_4, & F_5), which increases spectral power amplitude in the 2.4-3.6kHz region. The physiological cause of this effect is thought to be associated with a narrowing of the larynx tube (the space enclosed by the vocal folds, epiglottis, arytenoid cartilages and ary-epiglotic folds) combined with a widening of the pharynx such that the ratio of the larynx tube and pharynx cross sectional areas is 1:6 (Sundberg, 1974). The level or amplitude of the spectral power of the singer's formant has been found to vary depending on singer proficiency, vowel and phonation mode. However its centre frequency has been observed to be reasonably stable in the 2.4-3.6kHz region and invariant to vowel or phonation mode (Sundberg, 2001). Although the singer's formant has been studied acoustically based on the singer's formant frequency centre position and amplitude, little attention has been paid to the relevant formant frequency bandwidths in the singing voice.

The study of bandwidths is very important in the description of the singing acoustic spectrum as they can be interpreted as a measure of specific physiological aspects of voiced production. Whilst formant frequencies and amplitudes have been attributed to vocal tract area shape and air pressure, bandwidths can be attributed to vocal tract losses (Fant, 1972). Bandwidths however are difficult to accurately measure. One principal reason behind this is that the Fourier analysis of a vowel sound yields a set of harmonics spaced at frequency intervals equal to the fundamental frequency. Even with a male fundamental of 100Hz or lower, the harmonics are too far apart to accurately define the shape of the formant envelope. The use of Linear Prediction (LPC) (Markel & Gray, 1976) eases this situation somewhat, but the bandwidths extracted from the filter pole data are still considered unreliable.

This paper is presented as a pilot study, that aims observe the difference in formant bandwidths between spoken and sung vowels. It initially outlines mathematically how variant formant bandwidths can effect the level of the singer's formant. Then discusses an alternative method for the measurement of formant frequency bandwidths via a selective glottal reflection coefficient. It discusses data derived from a single male baritone classical singer to observe the formant bandwidths of a singer's formants.

Proceedings of the 9th Australian International Conference on Speech Science & Technology
Melbourne, December 2 to 5, 2002.
Accepted after full review

METHODS

Mathematical Discussion

Sundberg (2001) suggested a method for determining whether or not a sung vowel had a level of spectral energy that was greater than would be expected for that of a spoken vowel. Hence, would be likely to contain enhanced formant energy, such as is usually associated with a singer's formant. The method was based on the difference between the acoustical spectral energy of the first and third formants (L_3-L_1). Sundberg published predicted L_3-L_1 graphs for varying centre frequencies of the first formant (F_1) and gave guidance for a descriptive measure of the singer's formant. However Sundberg did not discuss the importance of the bandwidths of each formant or their role in shaping the final value of L_3-L_1.

A highly respected method for the calculation of the spectral envelope from formants and bandwidth data is Fant (1962). This method can be used to illustrate the effect of the modification of bandwidths on the L_3-L_1 value. Fant's method is based on the combination of four formants to form a spectral shape. Fant's equations have been a benchmark for formant bandwidth evaluation as they are not reliant on the LPC method as mentioned above. The equation set up is as follows:

$$L(f) = L_k(f) + L_{r4}(f) + \sum_{n=1}^{4} L_n(f) \text{ dB}$$

$L_k(f)$ =Glottal source and lip radiation
$L_{r4}(f)$ =Curve representing the higher formants
$L_n(f)$ =Curve representing each formant

Study	Vowel	Formant Frequency (Hz)				Formant Bandwidth (Hz)			
Story(1996)	/herd/	535.2	1572.9	2274.1	3072.9	56.2	121.8	164.5	279.1

Table 1: The formant frequencies and bandwidths for the American vowel /herd/ calculated directly from measured area functions using a sophisticated synthesis algorithm as published by Story.

By taking the example vowel in table 1 and varying the value of the first and third formant bandwidths, the effect on the resultant spectrum can be seen. These results are represented in figure 1.

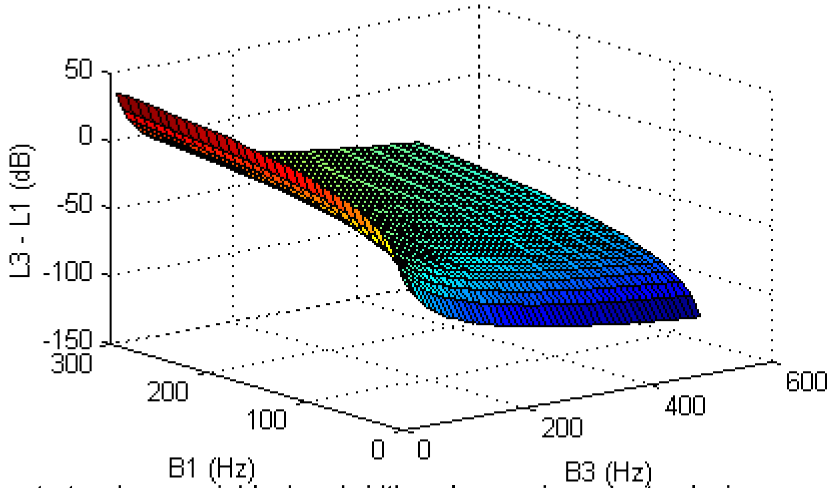

Figure 1: Modification of the bandwidth of F1 and F3 and its effect on the L3-L1. Quite clearly the maximum value occurs when B1 is maximised and B3 minimised.

The mathematical model above demonstrates how variable bandwidths play an important role in defining the spectral envelope of vowels. A defining characteristic such as the recognition of a singer's formant on the difference between L_3-L_1, must take into account the differing value of the bandwidths of each of the respective formants. A wide first formant in conjunction with a narrow third formant can contribute to a smaller L_3-L_1 value as was demonstrated mathematically in both models.

The Glottal Reflection Coefficient

As discussed previously, the evaluation of individual formant bandwidths is a difficult task. An alternative method that gives an indication of the nature of the average formant bandwidths can be determined from looking at a parameterisation of speech called the glottal reflection coefficient (GRC or μ_g). Reflection coefficients are a set of speech parameters (Markel & Gray, 1976) that provide a function of the acoustically reflective nature of individual sections of the vocal tract. The last coefficient or the coefficient responsible for indicating the reflective nature of the glottis is known as the glottal reflection coefficient. Due to the acoustic tube-like nature of the reflection coefficient model,

μ_g gives an indication of the losses of the entire vocal tract, as all losses are clustered at the glottis (Wakita & Gray, 1975, and Kasuya & Wakita, 1979). Kang (1976) demonstrated in synthetic simulations that as μ_g was increased towards unity, the bandwidths of the spectrum were reduced and the levels of the individual formants were increased. The μ_g can thus be thought of as a parameter of speech that delivers information on the average bandwidths (Wakita & Gray 1975) and its theoretical behaviour can be modelled as follows:

$$\mu_g = e^{\frac{-(\pi M B_{mean})}{f_s}}$$

μ_g = glottal reflection coefficient
M = number of reflection coefficients
B_{mean} = average pole bandwidths
f_s = sampling frequency

The Selective Glottal Reflection Coefficient

The μ_g gives us information about the average formant bandwidths from the perspective of the entire vocal spectrum. To obtain μ_g information dependant on a specific spectral region (ie the singer's formant region 2.4-3.6kHz), the selective linear predictive (SLP) method (Markel and Gray, 1976. Pg 150) was used to derive a selective glottal reflection coefficient (SGRC or μ_{sg}). The SLP algorithm is similar to the standard linear prediction algorithm only that a frequency specific spectral region is chosen to derive a series of reflection coefficients. The correct selection of the number of coefficients in the SLP is vital to give realistic values for μ_{sg}. For a sampling frequency of 11025Hz, the LPC size for the full frequency spectrum is set to 14. For the individual regions six coefficients were assigned for the 0-2kHz region and four for the 2.4-3.6kHz region. These coefficients were selected so that the frequency response of the selected region was as close as possible to the section of the full frequency spectrum that it was representing. If the coefficients were set too high then over-development of the spectrum occurred and the regional spectrum begins to resolve individual harmonics as formants. If too few coefficients are selected, then the region appeared under resolved and flat. As a general rule, it was found that the optimum number of regional coefficients is the number of full spectral coefficients multiplied by the fraction of the spectrum to be observed.

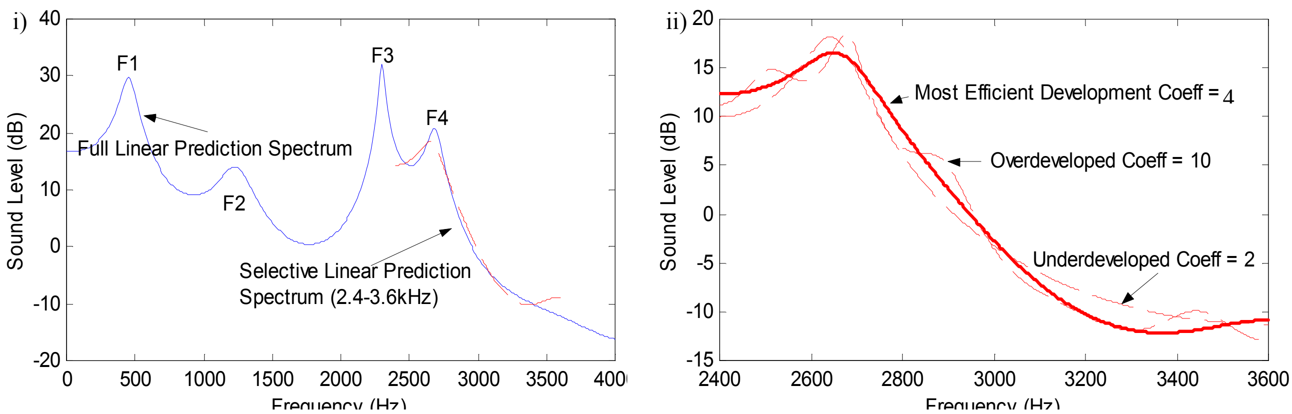

Figure 2. The use of Markel and Gray's Selective Linear Prediction. Figure 2i shows the frequency spectrum of a sung vowel /herd/ and the SLP resolution in the 2.4-3.6kHz region for six coefficients. Figure 2ii shows the SLP resolution in the 2.4-3.6kHz region using 2, 6, & 10 coefficients. A selection of 10 coefficients overdevelops the spectrum and begins to resolve harmonics, where as 2 coefficients under-resolves the spectrum and it appears flat.

Experimental Data

To observe the μ_{sg} and infer information about the formant bandwidths in the spoken and sung voice, acoustical information was recorded from a singer during speech and song. The singer (the first author) is not a professional opera singer, but his capability as a western classical opera singer has been demonstrated by him winning regional Australian championships in his voice category as a bass. The selected data was 5 repetitions of 11 steady state vowels recorded in spoken phonation and sung at $F_0 \approx 110$Hz (which was the approximate average speaking pitch). The 11 vowels chosen were randomly uttered for each sequence and were recorded in September 2001. The selected vowels were the Australian vowels in; /heed/, /hid/, head/, /had/, /hard/, /hod/, /hood/, hoard/, /hudd/, /who'd/, and /herd/.

Each vowel was then analysed, in turn, to determine: The long term average spectrum (LTAS) for the determination of the singer's formant level (L_3-L_1); The GRC (μ_g) for the entire spectrum; Two regional specific SGRCs (μ_{sg}) (a lower spectral region 0-2kHz and the singer's formant region 2.4-3.6kHz); and formant information from the LTAS. The results of which are presented in figures 3 i-v.

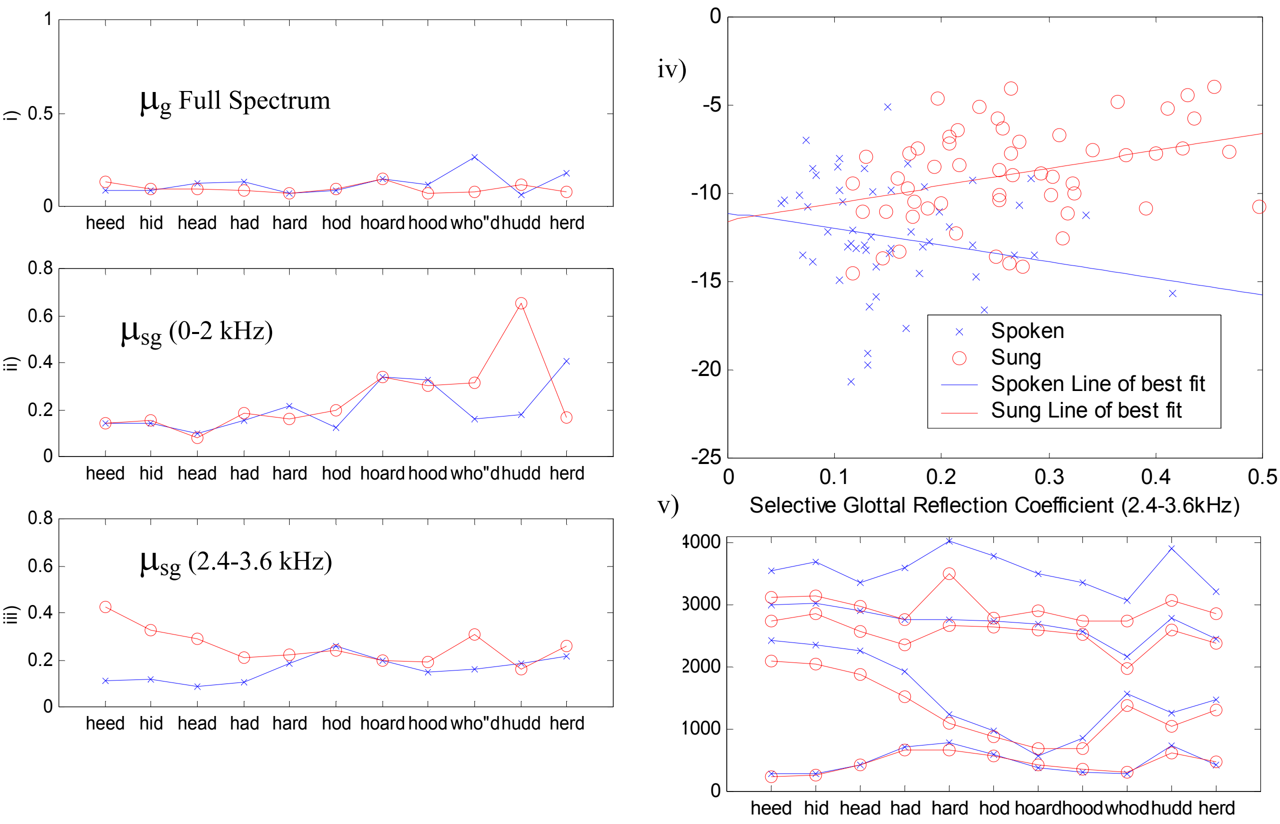

Figure 3(i-iii): Illustration of the value of the glottal reflection coefficient for both sung and spoken 11 vowels. The spoken vowels are represented by the /-x-/ line and the sung by the /-o-/ line. Figure i) is the complete spectral μ_g. Figure ii) is the lower spectral μ_g whilst Figure iii) is the singer's formant region μ_g. Figure 2(iv): Scattergraph of the μ_g(2.4-3.6kHz) and the L3-L1 value, with lines of best fit for the spoken and sung data. Figure 2(v): The first four formants for spoken and sung vowels for the data set.

RESULTS

The experimental results are displayed in five principal graphs. The first three (figures 3i-iii) illustrate the difference between glottal and selective glottal reflection coefficients in spoken and sung voice for the different vowels. Figure 3iv) is a scatter graph representing the L_3-L_1 value as a function of the μ_{sg} in the singer's formant region. Whilst figure 3v) is the formant tracks for the 11 vowels in both sung and spoken phonation.

The first three graphs display spoken and sung μ_g across the full spectrum and μ_{sg} for the 0-2kHz region and the 2.4-3.6kHz region, for all 11 vowels. Clearly there is little difference between spoken and sung μ_g across all vowels as shown in figure 3i. The spoken vowel /who'd/ did however display a slightly increased value for the μ_g. The second graph figure 3ii, was the μ_{sg} for the lower region between 0-2kHz. This μ_{sg} would give an average bandwidth figure for the first and second formant bandwidths. It shows that the front and back vowels have similar μ_{sg} values however the neutral vowels of /who'd/, /hudd/ and /herd/ have a dramatic difference for μ_{sg} in the 0-2kHz region. /who'd/ and /hudd/ have higher μ_{sg} values for singing whilst /herd/ has a higher μ_{sg} for speaking. Figure 3iii displayed the most observable difference in μ_{sg}. With the μ_{sg} observed in the singer's formant region (2.4-3.6kHz) there was quite an observable difference in the front and back vowels. All of the front vowels demonstrated in increase in μ_{sg} in sung data over spoken. The vowel /who'd/ displayed this effect also which could be due to the frontalised articulatory setting of the vowel /who'd/ in Australian

English. From observations of its formant pattern (figure 3v) it has a similar first and second formant positioning to other front vowels in this data set. The back and neutral vowels had similar μ_{sg} values, for both sung and spoken vowels.

Figure 3iv, is a scattergram plotting the values of L_3-L_1 (ie the difference in the spectral peaks of the third and first formants) against the μ_{sg} in the singer formant region. The spoken data represented by /x/'s does not show any specific correlation with the L_3-L_1 value as all points are clustered fairly centrally. The singing data represented by /o/'s does in fact show a subtle trend towards the fact that L_3-L_1 is in fact related to μ_{sg} (2.4-3.6kHz). A line of best fit for the singing data indicates that as μ_{sg} increases it is observed that the difference between first and third formants (L_3-L_1) also increases. Thus indicating that there may be a correlation between the μ_{sg} in the singer's formant region and the level of the singers formant reported by Sundberg (2001). When applying a line of best fit to the spoken data it had roughly the opposite effect. This line of best fit may however, have been influenced by outlying points. A visual inspection of the spoken points indicates that there is little to no correlation between μ_{sg} and the L_3-L_1 value, as they are all centrally clustered.

Figure 3v plots the formant tracks of the singing and spoken formant values taken for the 11 vowels. This displays similar formant tracks as observed by bass singers first reported by Sundberg (1974). The lowering of F_2, F_3, and F_4, can be clearly seen in the case of the singing data, and the clustering of the F_3 and F_4 can be observed at approximately 2.8-2.9kHz.

DISCUSSION

Discussion of Experimental Results

The experimental results showed some similarities with what was predicted from the original mathematical model. The first result came from the μ_{sg} calculations of the singer's formant region (2.4-3.6kHz) where for front vowels μ_{sg} was larger in singing than in speaking. Which would indicate a narrowing of the average bandwidths in this region. This ties in with the predictions from the mathematical theory, that indicated a narrow B_3 would increase the level of the singer's formant. The vowel /who'd/ also had a larger μ_{sg} value for singing but /who'd/ is a frontalised vowel in Australian English which would tie it in with the other normal front vowels. The back vowels demonstrate comparably similar values of μ_{sg} for sung and spoken vowels. Front vowels have a high tongue position that may be the reasoning behind the narrow bandwidths. Fant (1972) specifically states that a raising of the tongue against the palate could cause a narrowing of the bandwidths, however this should be replicated across all bandwidths and not be frequency dependent, as was observed with this data, where it is observed only in the singer's formant region.

To check for a relationship between the average bandwidths of the singer's formant region and the singer's formant level a scatter graph was plotted aligning the μ_{sg} (2.4-3.6kHz) with the L_3-L_1 level. All 110 samples of spoken and sung vowels were plotted, and a subtle correlation between an increasing μ_{sg} value and an increasing L_3-L_1 value for sung vowels could be observed. This result appears to support the original mathematical model. Which implied that the level of the singer's formant was seen to increase as a function of the narrowing of the third formant bandwidth. Which was observed indirectly by an increase in μ_{sg} (2.4-3.6kHz) implying a narrowing of the mean formant bandwidths in the singers formant region.

Articulatory speculation

As has been previously explained, that the glottal reflection coefficient provides an indication of the mean bandwidths of the formant spectrum. Bandwidths in their own right describe particular physiological traits of the vocal tract, and are thus are very important in the acoustical description of the singer's formant. Fant (1972) described the formant bandwidths as being directly influenced by four vocal tract loss contributors: cavity wall losses, friction, heat conduction and radiation losses. These losses are a function of either frequency, or area function of the vocal tract. The losses due to friction and heat conduction and the radiation losses are strongly dependent on the area function.

In the production of the singer's formant, the singing voice has been hypothesised as having: a narrow larynx tube region, followed by a largely open back pharynx region. (Sundberg, 1974). These two elements have not been seen to dramatically affect F_1 or F_2 but rather, lower the F_3, F_4 & F_5 and cluster these frequencies in the singer's formant range of 2.4-3.6kHz. The differences in area tract volume changes (from spoken language) would affect the bandwidth contributions of friction, heat and radiation losses according to Fant (1972). Fant indicated that the bandwidths of higher formants will become lower for the articulatory case by narrowing the lips or raising the tongue against the palate, and that this negative exponent adds to the cavity wall induced bandwidth increase towards low frequencies. This reflected a trend that was observed in our experimental data, with the increase of the μ_{sg} (2.4-3.6kHz), particularly in the front vowels which have the highest tongue posture. Fant's radiation losses were also particularly dependent on vocal tract area function, notably B_3, which was observed as having particular variability as a function of the maximum constriction from the glottis to the lips. If a singer had a particularly narrow larynx tube, this would no doubt add to the radiation loss. Fant (1972) also stated that B_3 could be narrowed in the typical case by narrowing the lips or raising the tongue against the palate which we identified previously as possibly effecting the μ_{sg} (2.4-3.6kHz) more so in front vowels than in back vowels.

Conceptually the reduction of the vocal tract losses in singing makes good pedagogical sense. A singer trains his/her voice to maximise efficient acoustical output. If a reduction in vocal tract losses increases acoustical output, then it follows that a singer should try to minimise the losses within the vocal tract. In situations where the singer is singing over an orchestra then the voice should seek to create as much acoustical energy as possible for aural differentiation purposes.

SUMMARY

This paper has outlined the effect of narrowing the formant frequency bandwidths on the acoustical spectrum of the singing voice and also its effect on the singer's formant. It has introduced the selective glottal reflection coefficient (μ_{sg}) and reported experiments with a single bass-baritone subject. A speculative articulatory correlation between the reduction of formant bandwidths based on vocal tract shape and acoustical losses was highlighted as a possible theoretical reason for this narrowing of specific bandwidths.

REFERENCES

Bloothooft, G & Plomp, R (1986) *The sound level of the singer's formant in professional singing*, J Acoust Soc Am 79: 2028-2033

Fant, G (1962) *Formant bandwidth data*, STL-QPSR 1/1962, 1 and 2.

Fant, G (1972) *Vocal Tract Wall Effects, Losses, and Resonance Bandwidths*, STL-QPSR 2-3/1972.

Kang, G. S. & Coulter, D. C. (1976) *600 BPS Voice Digitizer*, *ICASSP* 1976: 91-94.

Kasuya, H. & Wakita, H. (1979) *An approach to segmenting speech into vowel- and non vowel- like intervals.* IEEE trans. ASSP 27:319-327

Markel, J. & Gray, A. (1976) *Linear Prediction of Speech.* Springer-Verlag. New York.

Rothemberg, M (1982) *Acoustical interaction between the glottal source and the vocal tract*. In *Vocal Fold Physiology*, edited by K. N. Stevens.

Story. B, (1995) *Physiologically based speech simulation using an enhanced wave reflection model of the vocal tract.* PhD Disertation, Graduate College of the University of Iowa.

Sundberg, J. (1974) *Articulatory interpretation of the singing formant*, *J Acoust Soc Am* 55: 838-844

Sundberg. J, (2001) *Level and Center Frequency of the Singer's Formant*, *Journal of Voice* 15: 176-186

Wakita, H & Gray, A (1975) *Numerical determination of the lip impedance and vocal tract area functions.* IEEE trans. ASSP, 23: 574-580.