PERCEPTUAL CHARACTERISATION OF THE SINGER'S FORMANT REGION: A PRELIMINARY STUDY

Thomas J. Millhouse
Sydney Conservatorium,
University of Sydney,
Sydney, Australia
Thomas.Millhouse@bigpond.com

Frantz Clermont
JP French Associates
Forensic Speech and Acoustics Laboratory
York, United Kingdom
akustikfonetiks@yahoo.com.au

Abstract

This paper introduces the use of perceptual linear prediction (PLP) analysis to observe the nature of the singer's formant in the psychoacoustical domain. PLP analysis models the psychophysics of hearing to derive an estimate of the auditory spectrum that accounts for the known psychoacoustical theories of vowel perception. Building on the results obtained previously in Millhouse and Clermont (2004) this paper departs from previous studies on the acoustics of the singer's formant to discuss the nature of the singer's formant region within the perceptual subspace. The PLP technique demonstrates its robustness in the analysis of sung vowels at high pitches and identifies a perceptual singer's formant in soprano voices by representing the singer's formant as a continuous band of acoustic energy in the upper auditory region.

1. Introduction

Ever since the seminal study of Bartholomew (1934), singers, pedagogues and scientists alike have had a romantic affiliation with the notion of the 'Singer's Formant', i.e., a special formant that marks a clear difference between spoken and sung vowels in opera singers. The singer's formant is associated with 'good voice quality' in opera singing but, more specifically the perception of an opera singer's 'brilliance' (Vennard 1967), 'twang' (Yanagisawa, Estill, Kmucha & Leder 1989) or 'ring' (Ekholm, Papagiannis & Chagnon 1988). It is also hypothesised to be the psycho-acoustical basis for why an opera singer's voice is audible over a much louder orchestra (Sundberg 1974).

1.1 Acoustical nature of the singer's formant

From a classical discussion of speech science, the singer's formant is not a specific vocal tract resonance as such. It has been defined by Sundberg (1974) as a pronounced peak of acoustic energy located at about 3kHz (see left panel in Figure 1), which is formed by the clustering of the third, fourth and fifth vowel formants (F_3, F_4, and F_5). The articulatory basis for this effect is thought to be associated (see right panel in Figure 1) 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).

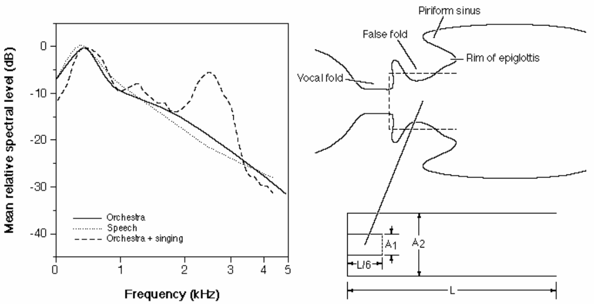

Figure 1**:** Singer's Formant: Acoustic-articulatory basis (Taken from Sundberg 1987.)

The amplitude of the spectral power of the singer's formant varies from singer to singer depending on their proficiency, voice type and phonation mode, but its centre frequency lies between 2.4 - 3.6 kHz invariant across a subject's vowel and vocal range (Sundberg, 2001).

The accurate classification of the singer's formant on an intra-vowel basis across all voice types remains an open research question due to the difficulty in measuring formant frequencies at high pitches (Vallabha & Tuller 2002). In the case of lower-pitched voice types (basses and baritones), acoustical formant analysis is relatively simple via linear prediction techniques (Clermont 2002). However, the wide spacing of harmonics in high-pitched singing has made the evaluation of acoustic formant frequencies problematic and unreliable for the higher-pitched voices. The characterisation of the singer's formant for these voice types (e.g., the Sopranos) remains therefore an active research question (Weiss 2001, Joliveau, Smith, & Wolfe 2004, and Barnes, Davis, Oates & Chapman 2004).

1.2 Perceptual nature of the singer's formant

Noted within the literature on the singer's formant is the presence, but not necessarily clearly defined nature, for both an acoustical and a pseudo-perceptual definition of this phenomenon. The scientist draws their conclusions on the singer's formant from the perspective of classical speech science, whereas the pedagogue adopts a more perceptually-oriented point of view. This dual approach has spawn differing interpretations of what is collectively known by both groups as the singer's formant. If the acoustical manifestation of the singer's formant is to be indeed associated with subjective pedagogical assessments such as 'good voice quality', then consideration of the perceptual aspects of this phenomenon must be undertaken. In view of the well-known association with perceptual impressions expressed as a singer's 'quality', 'brilliance' and 'ring', it is therefore proposed that the psycho-acoustical domain should afford a more complete characterisation of the singer's formant.

1.3 Exploratory bases in the psycho-acoustical domain

Two specific psychoacoustical observations present themselves as worthy of consideration in any perceptual interpretation of the singer's formant. They are the 'Effective Second Formant' theory of Fant & Risberg (1962) and the 'Spectral Peak Integration' theory of Chistovich, Sheikin & Lublinskaja (1978).

The effective second formant (F_2') theory arose from observations of vowels in the psychoacoustical domain. Fant

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

& Risberg (1962) indicated that the acoustical spectrum of the vowel could be reduced to two principal formants as perceptual cues for differentiating vowels. A perceptual first formant (F_1') strongly correlated with the acoustical first formant (F_1) and a second perceptual formant (F_2') that is a combination of the upper formant structure. In intelligibility studies of synthetic vowels, Fant & Risberg (1962) demonstrated that F_2' does not correspond to any particular formant, but is typically close to F_2 in simulating back vowels and lies between F_3 and F_4 for front vowels.

Chistovich et al (1978) 'Spectral Peak Integration' theory suggests that in the perception of speech, two spectral peaks are integrated into one when they are closer than a critical spectral distance $\delta_c = 3.5$ bark. This theory has specific implications for the singing voice, as the clustering of upper formants responsible for the singer's formant, could be seen as contributing to a single perceptual formant in the psychoacoustical domain.

1.4 Perceptually motivated studies of the sung vowel.

Perceptually motived studies of the singing voice have been undertaken and in general have had more success in defining spoken and sung vowels, then their pure acoustical counterparts. The study of Bloomploft & Plomp (1985) observed the differences in spoken and sung vowels using a 1/3 octave filter bank and a statistical analysis process known as Principal Component Analysis (PCA). This model delivered a two-parameter description of perceptual phonetic discrimination in the sung voice, based on the two highest principal components, which in turn were demonstrated to be representative of the acoustic formant frequencies, but did not address specifically the theories of Chistovich et al (1978) or Fant & Reisberg (1962).

Millhouse & Clermont (2004) used the analytical capabilities of perceptual linear predictive (PLP) analysis by Hermansky (1990) to estimate effective perceptual formants (F_1' and F_2') for spoken and sung vowels. An effective second formant F_2' was noticeable across all vowels in singing but not in spoken vowels. This differed from the perceptual studies of Fant & Risberg (1962) and the PLP work of Hermansky (1990) who identified that the effective second formant F_2' was only observed in front vowels for the spoken language. It was thus hypothesised that the lowering and clustering of the upper formant structure in sung vowels (attributed to the acoustical singer's formant), mostly likely contributed to the observable F_2'.

1.5 This study: Aims and outline

Following on from the work of Millhouse & Clermont (2004) this paper will turn specifically to the singer's formant region in the perceptual domain. This paper will consider three case studies. In the first case study (Section 3), we look closely at the acoustic and perceptual representations of singing and speech to identify the nature of the singer's formant in the perceptual domain. In the second case study (Section 4), we observe a Soprano's spoken and sung vowels at different pitches to address the performance of the PLP model for high pitched voices, and observe the effect of increasing pitch on the perceptual singer's formant region. This leads to specific insights into the intelligibility of sung vowels at high pitches and addresses the validity of the singer's formant in Soprano voices as it applies to the perceptual spectrum. Finally in the third case study (Section 5) we use acoustical and perceptual formant analyses of a Baritone subject to observe the contribution that the underlying acoustical formant structure plays in shaping the perceptual spectrum.

2. Methods and Materials

2.1 Materials

To observe the nature of the singer's formant in both the acoustic & auditory domains a combination of short vowel segments and longer sung phrases from popular operatic arias were collected for analysis. Two subjects were used, both of which are professional opera singers and native speakers of Australian English. In order to meet the requirement of a professional opera singer capable of producing a singer's formant, both subjects were chosen to satisfy a level of at least level four according to the Bunch & Chapman (2000) scale (see *Table 1*). Both subjects were recruited through the State Opera of South Australia and were either resident in South Australia or touring through South Australia at the time of recording. In addition to a spoken and sung aria performed by each subject, we also collected a complete set of Australian English vowels spoken and sung by each subject. Initially both subjects spoke five randomized tokens of 11 monosyllables in /hVd/ context, at their habitual speaking rate. Then queued with a series of pulsed sine wave tones, our subjects sang the same tokens at approximately the same pitch as the queued tone. The analogue signals were then sampled at 11,025 Hz and quantized to eight bits.

Voice Type	Age range (years)	Taxonomy	Native Country
Soprano	30-39	3.1b	Australia
Aria: 'Summer time' from Gershwin's *Porgy and Bess*			
Baritone	20-29	4.1b	Australia
Aria: 'Vecchia zimara senti' from Puccini's *La Boheme*			

Table 1**:** Voice type, taxonomy and aria selection of the two professional opera singing subjects.

2.2 Linear predictive formant extraction

To obtain the acoustic formant frequencies ($F_{1\text{-}4}$) an analysis-by-synthesis technique in accordance with Clermont (1992) was employed. Acoustic formants frequencies of the Baritone subject were extracted from the poles of the LP analysis through (Hanning) windowed frames of 30 msec duration, by steps of 10ms. For 10% of the spoken data the LP-order was increased to 16 from a default value of 14 and 20 for 20% of the sung data, in order to enhance the upper formant regions. For each steady-state section, the analysis-by-synthesis technique yielded four sets of formants for each sample vowel analysed. Due to the limitations of the LPC formant extraction process at high pitches only the baritones vowels were analysed to obtain their formant structure.

2.3 Perceptual linear predictive formant extraction

To obtain perceptual formant frequencies (F_1' and F_2'), a five coefficient perceptual linear prediction (PLP5) algorithm was employed. A full discussion of the PLP process is given in Hermansky (1990), but essentially it produces an auditory spectra using psychoacoustic transforms. The auditory spectra for our spoken and sung materials were thus defined on the bark scale by effectively band-pass filtering the corresponding waveforms through critical bands over the desired frequency range. These bands were then further processed using Hermansky's (1990) equal-loudness pre-emphasis and cubic-root amplitude compression prior to the derivation of autoregressive coefficients.

Once autoregressive coefficients had been determined for each vowel the perceptual formant frequencies can be solved by root solving of the PLP autoregressive coefficients on the

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

newly derived bark spectrum. This process required manual interpretation of the spectrograms as root solving becomes problematic when two peaks merge together in the auditory domain. According to Chistovich et al (1978), when two formants approach a bark separating of 3.5 bark, they will be perceptually integrated into the same peak. To ensure that merged formants were accurately represented by either a single peak, two merged peaks or two closely separated but specifically different formants, visual inspection and verification of the PLP5 log magnitude spectrum for each vowel was undertaken.

3. CASE ONE: Identification of the singer's formant in the perceptual and acoustic domain.

Bartholomew (1934) first identified what is now known as the singer's formant by looking at time frequency representations of sung vowels and recognising the pronounced spectral bands at approximately 2.8 kHz in male singers.

3.1 Acoustic spectrogram measurement and results.

Our first task was to derive time-frequency spectrograms of the male and female data to determine if a singer's formant was present in our subjects. Identification of the singer's formant in the Baritone subject was performed by visual inspection of the spectrograms for his sung and spoken arias (see Figure 2). The baritone subject's sung aria has a clearly observable singer's formant at approximately 2.7kHz that sets the spectra apart from the spoken equivalent.

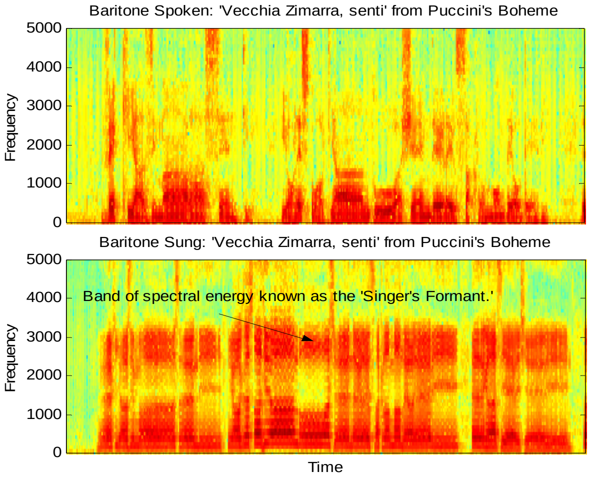

***Figure 2*:** Spectrograms of a Baritone's speech and singing indicating the singer's formant in the 3kHz region.

A clearly defined process for determining if a singer actually possesses a singer's formant remains unclear (Sundberg 2001). However, manual observation of the spectrogram remains a good rule of thumb for lower pitched male voices. This process fails however in the higher pitched voices, due to the wide separation of harmonics when singing at upper pitches. The inability to derive a formant structure due to the wide spacing of the harmonics, means that the singer's formant is unobservable (if indeed present at all) in the higher pitched voices. Figure 3 presents a spectrogram of our soprano's sung and spoken arias and it can be seen that for the sung aria the spectrum is dominated by the harmonic structure but little evidence is present of a detailed formant structure, unlike its spoken equivalent.

This case brings out elements of the debate that has arisen between the scientist and the pedagogue over whether or not a soprano possesses a singer's formant. The scientist describes the singer's formant as being present if there is a pronounced acoustical resonance at about 3 kHz that accounts for the 'brilliance', 'ring' and 'twang' of the singer's voice. For the soprano case illustrated here, there is no observable acoustic resonance and therefore no signer's formant, but the pedagogue would argue that there is still the perceptual observation of 'brilliance' and 'ring' in the soprano's voice.

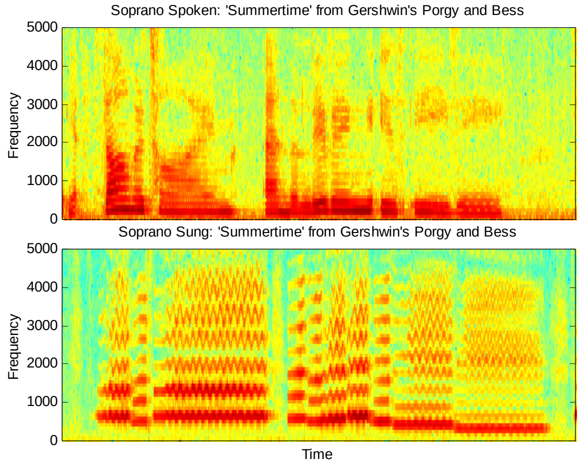

***Figure 3*:** Spectrograms of a Soprano's speech and singing in the acoustic domain.

3.2 Perceptual spectrogram measurement and results.

To determine whether the perceived 'brilliance' and 'ring' is explainable in acoustical and perceptual terms, the PLP5 algorithm was employed to generate spectrograms warped to the frequency characteristics of the human ear. Figure 4 provides such spectrographic representations of the same arias spoken and sung by the same soprano.

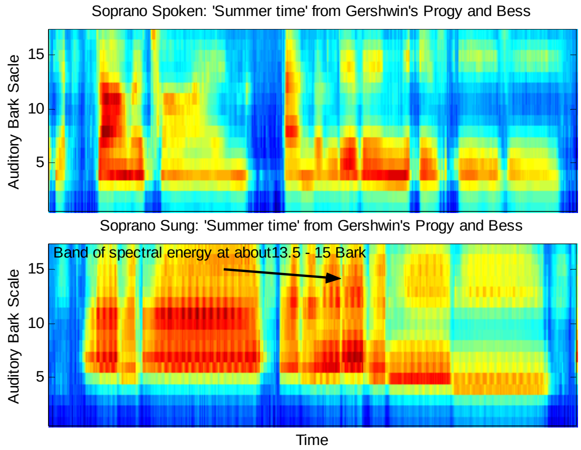

***Figure 4*:** Spectrograms of a Soprano's speech and singing in the perceptual domain.

Clearly noticeable in the sung PLP5 spectrogram is the presence of an additional band of spectral energy at approximately 13.5-14 Bark. This band of energy is not so clearly present in the spoken aria and appears to be a significant difference between the spoken and sung PLP5 spectrograms. This increased spectral brilliance could account for the perceived 'brilliance' and 'ring' in the soprano voice that is not so easily observed in traditional acoustic spectrograms. Of interesting note is that 13.5 Bark in the auditory frequency domain equates to approximately 3 kHz in the acoustic frequency domain. This increased spectral resonance area is essentially where we would have assumed the soprano's singer's formant to be located if present.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006. Copyright, Australian Speech Science & Technology Association Inc.

Accepted after full paper review

3.3 The presence of a soprano's singer's formant.

Considerable research has been undertaken to determine if a soprano actually possesses a singer's formant (Weiss 2001, Joliveau et al 2004, and Barnes et al 2004). Although a soprano can be described perceptually by the pedagogue as having a similar 'brilliant resonance', 'ring' or 'twang' as a bass or baritone, the lack of an easily observerable singer's formant in the acoustical domain has implied until recently that the soprano does not have a singer's formant.

The PLP5 derived perceptual spectrogram demonstrates though that a soprano's does in fact have a resonant band of energy that appears mostly independent of pitch and vowel in the perceptual domain. This observed perceptual singer's formant contains an resonance in the upper auditory spectrum at around 13.5-14 bark, similar in centre frequency to the singer's formant in the acoustic domain that most likely accounts for the 'brilliance', 'ring' and 'twang' in the soprano's voice.

Now that an auditory resonance seemingly independent of pitch and vowel has been observed our next task was to investigate the nature of this upper auditory resonance on an intra-vowel and variable pitch basis for the soprano voice.

4. CASE TWO: The effect of variable pitch on the vowel spectra of the Soprano's voice.

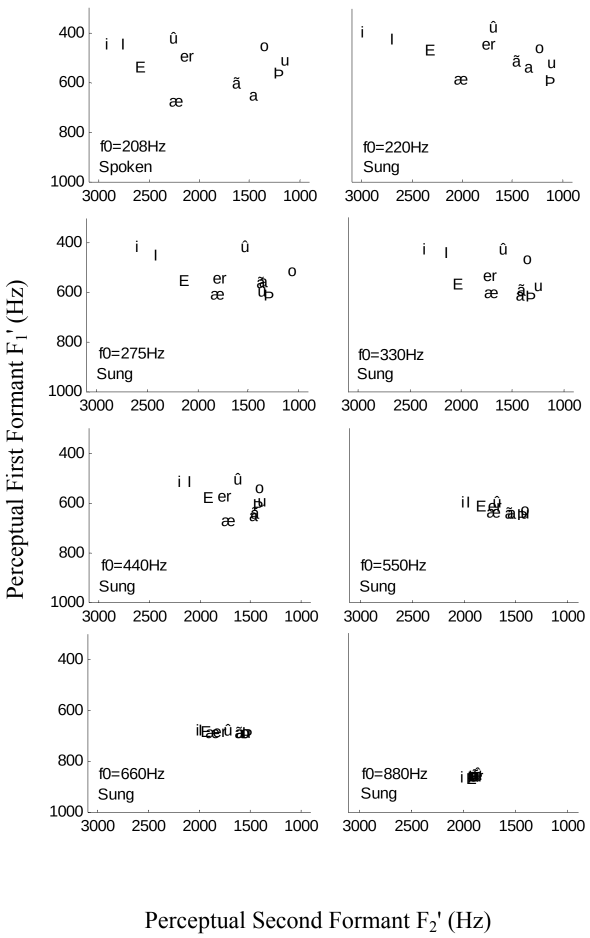

Figure 5: Soprano's perceptual formant values for the spoken and sung vowel /heed/. Where /i/ = heed, /I/ = hid, /E/ = head, /æ/ = had, /a/ = hard, /ɒ/ = hod, /u/ = hood, /o/ = hoard, /û/ = who'd, /ã/ = hudd, /er/ = herd.

Millhouse & Clermont (2004) identified that within the perceptual spectra of a baritone subject, a clearly noticeable F_2' was visible in all sung vowels, but F_2' was only visible within the front vowels for the spoken vowels. This initial observation implied that the acoustical singer's formant was manifesting itself as F_2' within the perceptual domain. To observe the nature of the perceptual formant movement as a function of pitch, a Soprano's vowel charts were generated and recorded at Figure 5. First, it is important to note that regardless of pitch the PLP5 process is capable of deriving perceptual formant values for each vowel. Second, similarly to the baritone subject, the soprano's back vowels are seen to have only a single perceptual formant when spoken, but two clearly defined perceptual formants when sung.

A first observation on the different pitch effects is that, for low f_0 the vowel distributions match closely the quasi-triangular shapes characteristic of the lower formant space. As pitch increases, F_1' is seen to increase to match the frequency of f_0. This could be due to the dominance that the f_0 has on the perceptual spectrum. The F_2' seems to vary depending on whether it is a front or back vowel. Front vowels observe a decrease in F_2' to finally cluster with F_1' at the highest observed frequencies, where as back vowels increase with F_1' prior to clustering with F_2' at high f_0. As pitch increases though, the PLP5 procedure infers that all vowels become somewhat spectrally homogeneous as they merge on the f_0.

The results obtained for the behaviour of perceptual formant frequencies as functions of pitch match closely the results reported in previous studies on vowel intelligibility at high pitches when sung by opera singer's (Benolken & Swanson 1990) and the principal-component results obtained in the studies of Bloomploft & Plomp (1985).

5. CASE THREE: The role of the underlying formant structure in the characterisation of the perceptual singer's formant.

The sets of results reported above suggest that there might be some connection between the second perceptual formant F_2' and the acoustical manifestation of the singer's formant. Indeed, F_2' has roughly the same bark frequency as the acoustic singer's formant and its presence appears almost independent of vowel. An intriguing question thus arises - Does the singer's formant manifest itself in the perceptual domain as F_2' and, if so, what roles does the acoustical formant structure play in contributing to this F_2'? As acoustic formant information can only reliably obtained for lower voice types, we now turn our attention to our baritone subject.

5.1 Acoustic and perceptual vowel formant charts.

Previously we explained that extracting formants at high fundamental frequencies is problematic, so observe the role that the underlying formant structure plays in establishing the perceptual formant structure only spoken and sung vowels from our Baritone subject are considered. Vowel sequence charts for the Baritone's acoustic and perceptual formants are detailed in Figure's 6 and 8. Figure 6 displays changes in acoustic and perceptual formant structure for each of the 11 spoken vowels. Our results replicate the perceptual studies of Fant & Risberg (1962), and of Hermansky (1990), in that within human perception and the PLP5 procedure up to two principal formants are derived for vowels in the auditory domain. Front vowels are observed to have two perceptual formants whilst back vowels only have one. The cases of /hoard/ and /hood/ are a special, in that despite being back vowels, a second perceptual formant was observed in the log-magnitude spectra, albeit increasingly small and barely observable from the PLP5 spectrogram. Most likely, this was due in part to the particularly low centre frequencies and

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

small separation distances between F_3 and F_4, and the low and merging F_1 and F_2.

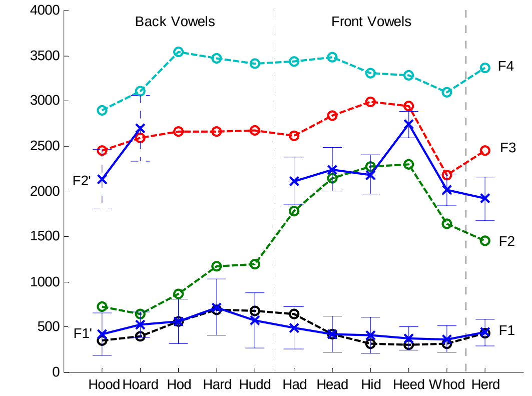

***Figure 6*:** Acoustic and Perceptual Formant patterns for spoken Australian English in /hVd/ context.

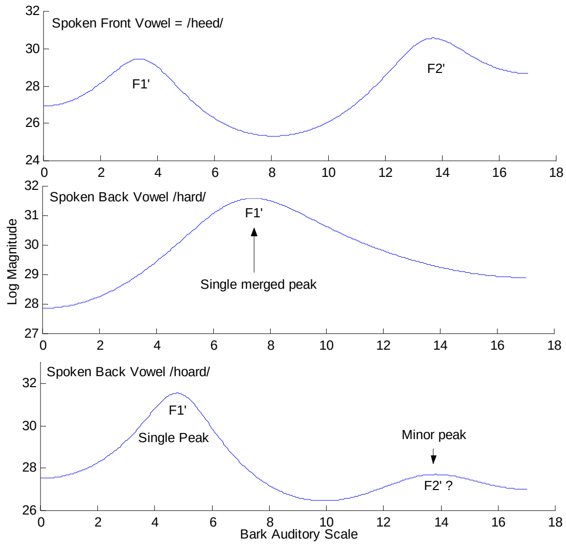

***Figure 7*:** Perceptual vowel spectra for spoken /heed/, /hard/ and /hoard/ indicating the small but defined F_2'

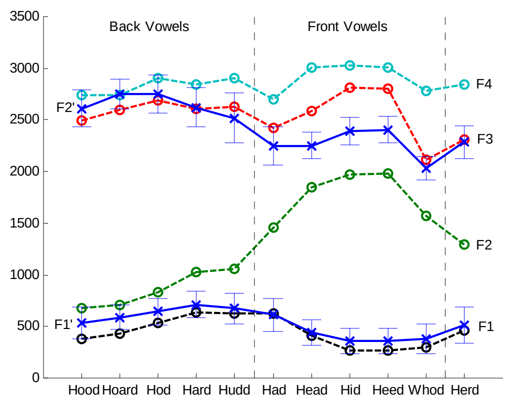

Figure 8: Acoustic and Perceptual Formant patterns for sung vowels at 110Hz in the Australian English in /hVd/ context.

Whether or not this smaller F_2' is significant with regards to perception is unknown but considered unlikely due to the phonetic discrimination in previous studies of Australian English back and front vowels. More than likely it is a manifestation of the speaker-dependent characteristics of the upper formant spectrum. Figure 7 demonstrates the relative magnitude of this small but appreciable perceptual formant in the baritone's spoken vowels /hoard/ and /hood/.

Figure 8 is the same classic vowel sequence chart as which is given in Figure 6 but this time is detailed for the sung vowels. The acoustic formants are generally lower due to the lengthened vocal tract and the clustering of the upper formants around 2.8kHz, as expected for the male sung vowels. These results replicate the studies of Clermont (2002) and Sundberg (1970).

The perceptual formants also differ from their spoken counterparts in that the sung vowels have a defined F_2' for all vowels. This F_2' appears to be located between F_2 and F_3 for the sung front vowels (as with the spoken equivalents) but is then also present for back vowels located between F_3 and F_4. Direct observations of the both sung and spoken perceptual vowels would indicate that their mean centre frequencies are approximately equal for each vowel, but that the bandwidths of each formant appear to be wider in the spoken case than the sung. This is especially apparent in the observation of the F_2' for the spoken /hoard/ and /hood/ where due to its significantly reduced amplitude the autoregressive modelling has afforded it a much wider bandwidth.

5.2 Correlation of F_2' with the singer's formant

To observe the influence that each acoustic formant has on F_2', a Multiple Linear Regression (MLR) technique with a resultant Pearson correlation coefficient (*r*) is used. The *r* determines the correlation of the PLP derived F_2' and the MLR derived F_2'. By using the MLR technique, F_2' can be determined for any single or combination of formants, and their contribution of these formants to the PLP F_2' can be determined by observing the Δr from the formant complete MLR F_2' to the format depreciated F_2'. The bandwidths are included in this process as they dictate spectral shape and have been demonstrated by Millhouse, Clermont & Davis (2002) to influence the singer's formant. Initially all formants and bandwidths are used to determine the complete MLR algorithm of the form in Eq. 1, and its resultant *r*.

$$F_2' = a_0 + a_1F_1 + a_2F_2 + a_3F_3 + a_4F_4 + a_5B_1 + a_6B_2 + a_7B_3 + a_8B_4 \quad \text{[Eq. 1]}$$

By removing individual formants and their respective bandwidths from the MLR equation and recalculating *r*, the reduction in the new *r* from the *r* calculated with a full format compliment, indicates the contribution of the formants to F_2' in the reduced MLR equation. The smaller the Δr the more predominate the formants in the reduced MLR equation have on F_2'. Table 2 details the *r* for each combination of formants and bandwidths. It can been seen that whilst the highest *r* values always occur when all formants and bandwidths are used in the MLR equation, there are significant groupings of formants that seem to have a predominate influence on the centre frequency of F_2' because their isolated use in the MLR equations only causes a slight deviation in the value of *r*.

Rank	Back vowels		Front vowels	
	F_x , B_y	*r*	F_x , B_y	*r*
1	$F_{1\text{-}4}$, $B_{1\text{-}4}$	0.87691	$F_{1\text{-}4}$, $B_{1\text{-}4}$	0.97159
2	$F_{2\text{-}4}$, $B_{2\text{-}4}$	0.86104	$F_{1\text{-}3}$, $B_{1\text{-}3}$	0.96847
3	$F_{3\text{-}4}$, $B_{3\text{-}4}$	0.84828	$F_{2\text{-}4}$, $B_{2\text{-}4}$	0.96261
4	$F_{1\text{-}3}$, $B_{1\text{-}3}$	0.82128	$F_{2\text{-}3}$, $B_{2\text{-}3}$	0.95670
5	$F_{2\text{-}3}$, $B_{2\text{-}3}$	0.80577	$F_{3\text{-}4}$, $B_{3\text{-}4}$	0.94214
6	$F_{1\text{-}2}$, $B_{1\text{-}2}$	0.70314	$F_{1\text{-}2}$, $B_{1\text{-}2}$	0.93009

***Table 2*:** Stepped multiple linear regression and resultant Pearson coefficients.

For the back vowels the F_3 and F_4 are seen to be the grouping of two formants that have the most influence on F_2'.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006. Copyright, Australian Speech Science & Technology Association Inc.

Accepted after full paper review

Whilst rated third in *r* value they are close in *r* value to the complete MLR and are also included as primary contributors in the second rated combination of F_2, F_3 and F_4. This agrees with our Figure 8 in that for back vowels F_2' lies between F_3 and F_4. For front vowels, we see that F_2 and F_3 seem to play the more important role due to their high *r* values and inherent inclusion in both groups of three formants that are rated higher than they are. Again, this agrees with our Figure 8 as F_2' is seen to lie between F_2 and F_3.

6. Conclusion

The purpose of this paper was to introduce the concept of the singer's formant as defined within a perceptual context. The perceptual singer's format has been demonstrated to be a continuous resonance located within the upper auditory band. This resonance is mostly independent of pitch and vowel, and is observed easily in both our soprano and baritone subject. This interpretation departs from previous acoustic interpretations of the singer's formant in that observations of the singer's formant in the perceptual domain accounts more fully for the pedagogical observations of 'ring', 'brilliance' and 'twang' that are observed in the opera singer's voice. The perceptual singer's formant has been shown to manifest itself as the perceptual second formant F_2' and has been shown to be contributed by the underlying acoustic formants F_2, F_3 and F_4 of varying degrees depending on the vowel.

This is a preliminary study, and as such considerable more work needs to be conducted before a more complete interpretation can be stated. That being said it is the first study of its kind that specifically attempts to characterise the singer's formant by linking the interaction of the acoustic and perceptual domains. Results in this study are promising but clearly more subjects of varying voice types need to be analysed to ensure that they fit within this characterised model of the perceptual singer's formant.

The preliminary results of this study do demonstrate and link the concept that the singer's formant is a phenomenon not just of acoustics, but also of human perception. This concept fits well with the pedagogical hypothesis that the operatic singing voice has evolved over the centuries to match a perceptual concept of vocal beauty.

7. Acknowledgements.

Both authors would like to acknowledge the assistance of A/Prof Dianna Kenny at the Australian Centre for Applied Research in Music Performance at the Sydney Conservatorium of Music, University of Sydney for her valued support in the production of this paper.

8. References

Barnes, J. Davis, P. Oates, J. & Chapman, J. (2004). The relationship between professional operatic soprano voice and high range spectral energy. *J. Acoust. Soc. Am.* 116(1), 530-538.

Bartholomew, W. (1934). A physical definition of good voice quality in the male voice. *J. Acoust. Soc. Am.* 6, 25-33.

Benolken, M. S. & Swanson, C. E. (1990). The effect of pitch-related changes on the perception of sung vowels. *J. Acoust. Soc. Am.* 87(4), 1781-1785.

Bloomploft, G. & Plomp, R. (1985). Spectral analysis of sung vowels. II. The effect of fundamental frequency on vowel spectra. *J. Acoust Soc Am* 77, 1580-1588.

Bunch, M. and Chapman, J. (2000). Taxonomy of Singers Used as Subjects in Scientific Research. *J. Voice*. 14(3), Pg 363-369.

Chistovich, L. A., Sheikin, R. L., and Lublinskaja, V. V. (1978). Centres of gravity and spectral peaks as the determinates of vowel quality. In Lindblom, B. & Ohman, S. (Eds) *Frontiers of Speech Communication Research*, (143-157) New York: Academic.

Clermont, F. (1992). Formant-Contour parameterisation of vocalic sounds by temporally-constrained spectral matching. *Proc. of the 4th Australian Conference on Speech Science and Technology* 48-53.

Clermont, F. (2002). Systemic comparison of spoken and sung vowels in formant-frequency space. *Proc. 9th Australian International Conference on Speech Science and Technology.* 124-129.

Ekholm, E., Papagiannis, G., & Chagnon, F. (1988). Relating objective measurements to expert evaluation of voice quality in Western classicial singing: Critical perceptual parameters. *J. Voice*. 12(2), 182-196.

Fant, G. & Risberg, A. (1962). Auditory matching of vowels with two formant synthetic sounds. *STL-QPRS* 4:7-11.

Hermansky, H. (1990). Perceptual linear predictive (PLP) analysis of speech. *J. Acoust. Soc. Am.* 87(4).

Joliveau, E., Smith, J., & Wolfe, J. (2004). Vocal tract resonances in singing: The soprano. *J. Acoust. Soc. Am* 116(4), 2434-2439

Millhouse, T. J., Clermont, F., & Davis, P. (2002). Exploring the importance of formant bandwidths in the production of the singer's formant. *Proc. of the 9th Australian Conference on Speech Science and Technology* 373-378.

Millhouse, T. J., & Clermont, F. (2004). Systematic comparison of spoken and sung vowels in the acoustic-auditory domain. *Proc. of the 10th Australian Conference on Speech Science and Technology.*

Sundberg, J. (1970). Formant structure and articulation of spoken and sung vowels. *Folia Phoniatrica,* 22, 28-48.

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

Sundberg, J. (1987). *The Science of the Singing Voice* Northern Illinois: University Press.

Sundberg, J. (2001). Level and Centre Frequency of the Singer's Formant. *J. Voice.* 15(2) 176-186.

Vallabha, G. K. and Tuller, B. (2002), Systematic errors in the formant analysis of steady-state vowels. *Speech Communication.* 38, 141-160.

Vennard, W. (1967). *Singing – the mechanism and the technic* (rev. ed.). New York: Carl Fischer.

Weiss, R., Brown, W. S., & Morris, J. (2001). Singer's Formant in Sopranos: Fact or Fiction? *J. Voice*, 15(4) 457-468.

Yanagisawa, E., Estill, J., Kmucha, S. T., & Leder, S. B. (1989). The contribution of aryepiglottic constriction to 'ringing' voice quality – a videolaryngoscopic study with acoustic analysis. *J. Voice*, 4, 342-350.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review