Vowel placement during operatic singing: '*Come si parla*' or '*aggiustamento*'?

Thomas J. Millhouse & Dianna T. Kenny

Australian Centre for Applied Research in Music Performance, Sydney Conservatorium of Music, University of Sydney, Australia

`thomas.millhouse@gmail.com, d.kenny@usyd.edu.au`

Abstract

This study explored two tenets of the Italian *Bel Canto* operatic singing technique - "*Come si parla" and* "*aggiustamento.*" Articulatory changes in the lower formant vowel space of 11 spoken and sung vowels were systematically examined in six male singers. Results showed that singers influence the placement of the lowest formant frequencies in the sung vowel space using both a lowered larynx and modified vowel articulation (*aggiustamento)* with rising pitch, especially above 220Hz.

Index Terms: singing voice, formant structure, formant tuning, sung vowels.

1. Introduction

The teaching principles of the classical Italian *Bel Canto* school of singing emblemized by the aphorism "*Come si parla*" - "as one speaks" [1], form the theoretical underpinning of this paper, which examines the buccopharyngeal placement of vowels during singing. This *Bel Canto* principle asserts that for the same speaker, the tongue and upper pharyngeal postures used in the sung vowels should have the same placement as that of their spoken vowels. However, the same school of singing also encourages "*aggiustamento*" or vowel modification during singing [1], [2]. This process involves adjustments to tongue placement and shape within the vowel in the singer's upper registers in order to maintain a vocal scale that is considered pure, unbroken and uninterrupted. Presumably, this adjustment should constitute the smallest possible change compared with vowel placements in the lower registers, which is nonetheless sufficient to achieve the desired vocal quality in operatic singing. This study explored whether operatic singers trained in the *Bel Canto* method apply gentle upper pharyngeal vowel modifications – *aggiustamento* to achieve a uniform scale.

We examined the resonant formant patterns of spoken and sung vowels at varying pitches in six male singers. An established method in speech science, resonant formant frequency analysis identifies vocal tract settings in spoken vowels. However, the method is more difficult to apply in the measurement of the sung formant structure using traditional formant tracking techniques [3], because of the complexities of sung vowels that are not shared with their spoken equivalents. For example, a singer's vibrato superimposes a sinusoidal harmonic movement on the vowel and the widely spaced harmonics at higher pitches make vowel formants inherently difficult to measure in singing. Despite these difficulties, comparative studies of spoken and sung vowels [4],[5] and [6] have been instructive, but have tended to focus on individual vowels and their subsequent formant patterns. Thus any articulatory interpretations would subsequently be focused on a single vowel rather than the vowel space as a whole. Our study applies a similar approach undertaken by [7], where the systemic modifications of the vowel system as a whole is analysed to identify the presence of common articulatory gestures needed to produce the sung vowel at various locations in the vocal range.

2. Material and methods

2.1. Participants

Six male singers of differing vocal types and proficiency were selected to provide spoken and sung data for this study. All singers were professional operatic artists engaged by the State Opera of South Australia at the time of recording and were all rated above level three according to the Bunch and Chapman [8] Voice Taxonomy scale. In this scale Level 3.1 or 4.1 indicates that the subject is a professional opera singer at the "National/Big City" (3.1) or "Regional/Touring" (4.1) level of proficiency. The sub classifications a, b or c differentiate between Major Principal (a), Minor Principal (b) or Chorus artist (c). As indicated by Table 1 our subjects comprised three tenors, one baritone and two basses, with a mixture of age and professional experience. The ratings recorded in Table 1 provide a measure of professional experience which can in turn give a measure of the reliability of the singer's to produce the required sung vowels consistently using the Bel Canto operatic technique.

Table 1: *Age, voice type and taxonomic classification of participant singers.*

Singer	Age range (years)	Voice Type	Taxonomy
AL	27	Tenor	4.1c
JP	26	Tenor	4.1b
EE	47	Tenor	3.1b / 4.1a
TM	27	Baritone	4.1b
DG	33	Bass	4.1c
PA	28	Bass	3.1b / 4.1a

2.2. The database of spoken and sung vowels

Data comprised 11 steady state vowels spoken and then sung at various pitches selected according to the singer's voice type. The choice of pitch was based on the 'A' major scale with the universal constant of A4 = 440 Hz as the anchor pitch. Initially, participants each spoke five randomised tokens of 11 monosyllables ("heed", "hid", "head", "had", "hard", "hudd", "hod", "hoard", "hood", "who'd" and "herd") at their habitual speaking rate. These stimuli were then queued with a sinusoidal tone at pitches appropriate to the singer's vocal range and the singer sang each vowel at the required pitch with three repetitions. Typically, tenors sang between 110 Hz – 440 Hz and the baritone and basses sang between 68.75 – 330Hz. During the recordings, a phonetically trained singer and pedagogue was present to

Accepted after peer review of full paper

provide coaching and vowel correction for each of the singers, to compensate for incorrectly pronounced vowels that occurred during recording.

Each recorded vowel was segmented and independently verified by the same phonetically trained singer from the recordings on a vowel by vowel basis. Incorrectly pronounced vowels, not identified during the recordings were removed from the data subset and the remaining isolated vowels were then sampled at 11,025 Hz and quantised to 8 bits before being catalogued for analysis.

2.3. Formant Frequency Measurement

The analysis of sung vowels is challenging because formant frequency measurements become problematic when the pitch (or fundamental frequency) exceeds 250Hz [3]. This is the approximate pitch at which the intelligibility of sung vowels in isolated context also degrades [9]. For this study formant frequency measurement was performed within the pitch limitations of linear predictive (LP) analysis [3] while remaining sensitive to the perceptual limits of intelligibility as reported for the singing voice in [9]. In addition to LP analysis, intra- and inter-vowel formant statistical methodologies for error checking provided an additional measure of quality for the predicted formant frequencies.

For each vowel the steady state voiced section of every vowel nucleus was isolated spectrographically prior to formant frequency measurement. Once the steady state section was obtained, formant frequency estimation was conducted via LP methods. For each steady-state section the linear predictive analysis yielded a set of frame-by-frame poles among which F_1 and F_2 were estimated using cepstral analysis-by-synthesis (AbS) and dynamic programming (DP) algorithms outlined at [10].

2.4. Inter-Frame Variability

Inter-frame variability over the steady state segment of the vowel is a measure of formant variability from the mean through the segment. It provides a measure of reliability of the formant estimation technique during the presence of vibrato for sung vowels. Vibrato appears as an undulating sinusoidal variation in pitch and amplitude that can affect the formant centre frequencies and the predictive capability of any formant tracker. To minimise the effects of vibrato the formant centre frequencies were averaged across the steady state selection of the vowel and the standard deviation of the intra-vowel variance was used as a measure of reliability.

Inter-frame variability also identifies when formant frequency estimation has been corrupted by high pitched phonation. As pitch increases, formant frequencies predicted via LP begin to merge on a harmonic frequency [3]. Thus, in our study when a frame-by-frame formant frequency appears to be modulating at the same rate as the predicted pitch we can immediately dismiss this sample as corrupted due to the pitch limitations of the LP formant estimation procedure.

2.5. Inter-Token Variability

Averaged vowel formant patterns exhibit variability caused not only by the limitations of the formant frequency measurement techniques, but also by singers' intrinsic variability when producing vowels spoken and sung in the same manner. Analysis of the variability of average vowel formants from vowel to vowel provides information regarding the variability within a singer's pronunciation for both singing and speaking.

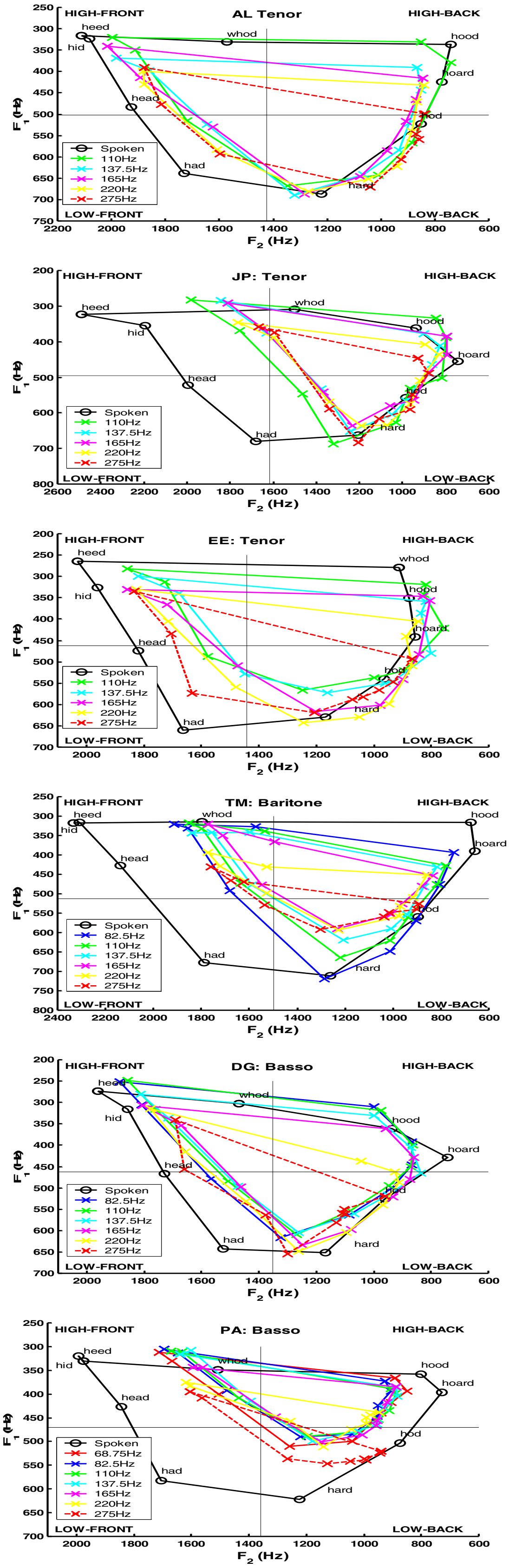
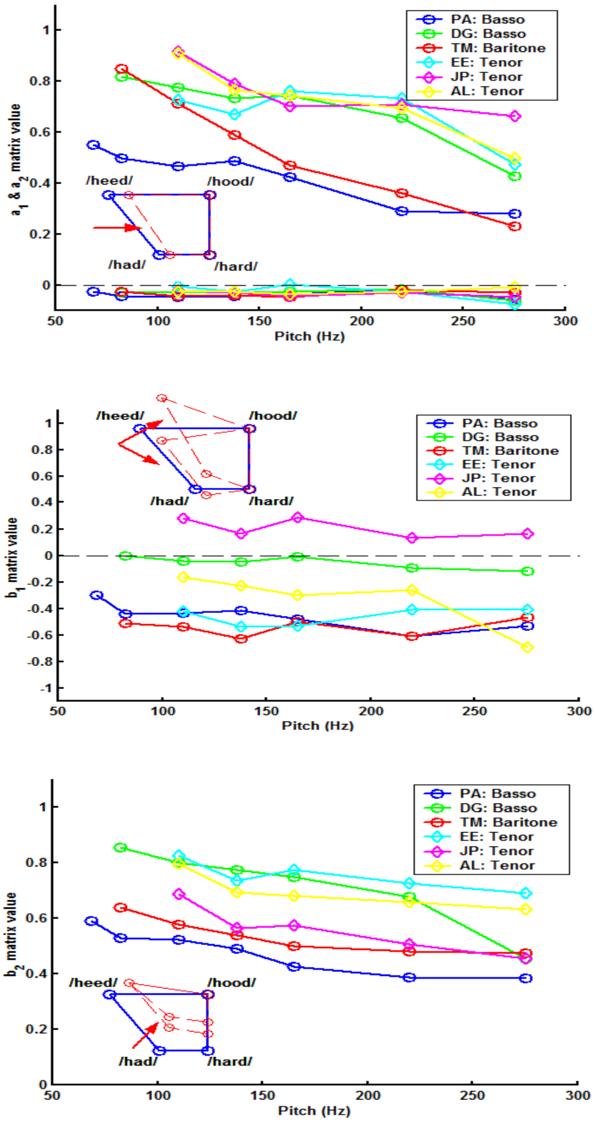


Figure 1: *Planer distributions of F_2 vs F_1 plotted for each singer. AL, JP, EE, TM, DG, PA.*

3. Results

Our six singers' results provided in Figure 1 present good examples of the general trends previously reported differentiating the lower formant patterns of spoken and sung vowels [7],[11] and [12] . Comparisons between spoken and sung vowels for the two lowest formant frequencies indicate that there is very little difference for F_1; however, for F_2 we note a marked dissimilarity between the front and back vowels in their spoken and sung counterparts. F_2 decreases significantly for the front vowels but is generally maintained for the back vowels, indicating an asymmetric shift from front to back vowel configuration in the sung vowels. The effect that rising pitch has on the vowel space is more dramatic. The increase in pitch can be observed to compress the vowel space towards the lower back quadrant. Noticeably, these formant transitions are not uniform across all singers' and pitch variations. Each singer presents speaker / singer specific effects that warrant further investigation and discussion.



Figures 2a-c: *Representations of the matrix parametrics 'a' and 'b' as a function of rising pitch. Inset describes the resultant effect of parametrics 'a' and 'b' on the vowel space. Blue represent spoken; red dash represent sung vowels.*

4. Discussion

4.1. Phonetic variability in F_1-F_2 space

Previous research on systematic transformations of the spoken and sung vowel space demonstrated that the asymmetric transformation supported the hypothesis that the lowered larynx position dominated the spoken-to-sung transformation in bass voices [7]. However, these studies used only three bass singers and did not consider pitch variation. This study extended the findings of previous research using systematic transformation technique to explore this phenomenon in tenors, baritones and basses. In addition, we examined whether this technique provided additional information regarding the variability of acoustical settings that dominate the differences in spoken and sung vowels.

Taking the systematic transformations reported in [7] and incorporating an additional dimension of pitch *'n'* our systematic transformations are represented in the equation (1)

$$\begin{bmatrix}\tilde{F}_{1,n}\\ \tilde{F}_{2,n}\end{bmatrix}=\begin{bmatrix}a_{0,n}\\ b_{0,n}\end{bmatrix}+\begin{bmatrix}a_{1,n} & a_{2,n}\\ b_{1,n} & b_{2,n}\end{bmatrix}\times\begin{bmatrix}F_1\\ F_2\end{bmatrix} \qquad (1)$$

such that each of the derived formant frequencies are a transformation around a mean term a_0 or b_0 and a weighted function of F_1 and F_2 for each pitch *'n'*. These '*a*' and '*b*' parametrics plotted against rising pitch provide an indication of the trends in the phonetic space and articulatory setting as a function of rising pitch.

Figures 2a displays the parametric disposition for $a_{1,n}$ and $a_{2,n}$. For all singers there was a common $a_{2,n} \approx 0$ and a value for $a_{1,n}$ varying between 0.4 and 1. All the singers also presented a gradual decrease in the $a_{1,n}$ parametric. This indicates a compression of the phonetic space polygon from the front vowels towards the back vowels. Inset into Figure 2a presents a diagrammatic representation of the compression of the front vowels towards the back vowels as a function of an $a_{1,n}$ varying between 1 and 0.6 and $a_{2,n} \approx 0$. Soloists rated as an 'a' or 'b' in terms of their professional experience (PA, TM) are also the singers who have the greatest compression of the vowel space towards the back vowels. In contrast, the chorus artists (DG & AL) tended to maintain their sung vowels as per their spoken vowels except at the higher pitches of their range.

The results for the '*b*' parametrics are more variable compared with '*a*'; the most variability between singers occurred as a function of both voice type and voice taxonomy. Parametric *'b_1'* shown in Figure 2b represents a raising or lowering of the sung front vowels into the jaw high or low configurations of the vowel space. (Inset to Figure 2b demonstrates this transformation.) In this example general trends of voice type of taxonomy are not so readily observable but tentatively it could be noted that the soloists (PA, TM & EE) had lower and negative *'b_1'* vowel shifts indicating a shift towards the low configurations; the chorus artists (AL & DG) showed less vowel shifting. The singer JP, a tenor, had a positive *'b_1'* coefficient that indicates a lifting of the vowel space into the jaw high configuration.

The final parametric *'b_2'* (Figure 2c) represents a compression of the jaw lower configured vowels (/had/, /hard/ & /hudd/) towards the jaw higher configuration. This setting appears to be reasonably uniform for all singers varying in value between 0.4 and 0.8 with a slight decrease as a result of rising pitch. Once again soloists (PA, TM & EE) have lower values indicating greater compression than chorus artists (AL & DG).

4.2. Vocal tract lengthening

The primary mechanism for articulatory shifts between spoken and sung vowels described in [7] was vocal tract lengthening. This study confirmed that vocal tract lengthening is a significant articulatory gesture, however other more subtle upper pharyngeal modifications may also be occurring, especially as the singer reaches the upper limits of their pitch range. Notably, the parametric results presented by *'a'* and *'b'* indicate that vocal tract lengthening is not uniform across vowels and pitch. The non-zero result for parametric 'b_1' presents evidence against the sole influence of vocal tract length. As a uniform matrix would be represented by the uniform matrix multiplied by a constant as shown in equation (2):

$$\begin{bmatrix} \widetilde{F}_{1,n} \\ \widetilde{F}_{2,n} \end{bmatrix} = \begin{bmatrix} a_{0,n} \\ b_{0,n} \end{bmatrix} + \psi_n \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \qquad (2)$$

where the symbol Ψ_n represents a constant of vocal tract lengthening as a function of pitch *'n'*. As clearly our *'b_1'* is non-zero it would be prudent to infer that vocal tract lengthening is not the only systemic articulatory setting that differentiates spoken from sung vowels.

4.3. Formant Tuning

Formant tuning has been described and observed in sopranos [5], [6] and tenor voices [4]. Formant tuning is a phenomenon whereby formant frequencies are slightly modified by the upper pharyngeal articulators to match the underlying harmonic structure of the voice source. For the higher ranges of sopranos F_1 in particular is matched to the fundamental frequency. This is typically observed on back vowels sung at a fundamental that exceeds the normal F_1 of the vowel. Formant tuning at higher frequencies may account for the movement of the vowel space towards the low jaw configuration as observed in Figure 1 for the higher fundamental frequencies. However, the effect of formant tuning has only been noted in tenors and soprano voices at pitches higher than our maximum analytical values of 275Hz. Figure 3 provides evidence of formant tuning in our bass subject PA.

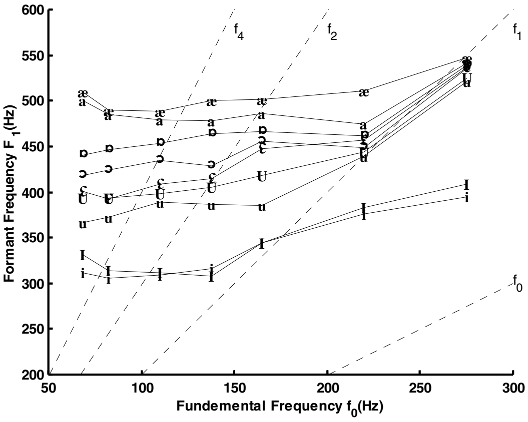


Figure 3*: F_1 vs f_0 for the Bass subject PA indicating the formant tuning of F_1 to the second harmonic f_1 at the higher pitches of this subjects range.*

Unlike the sopranos and tenors, this bass appears to tune F_1 to the second harmonic f_1 not the fundamental frequency f_0. This tuning occurs for all vowels except /heed/ and /hid/ which do increase with the fundamental but lie at a frequency roughly between the first and second harmonics. In this case the subject might wish to preserve the intelligibility of the high front vowels. For the remaining vowels the subject appears to be matching his F_1 to the second harmonic to maximize loudness and vocal efficiency for these vowels in a manner similar to soprano voices in [5]. In our participants *aggiustamento* was employed (see Figure 1) mostly at pitches above 220Hz. This frequency could indicate that this is a critical transition point in the male voice at which modifications are required in the upper-pharyngeal tract to preserve the quality of timbre in the voice.

5. Conclusions

This paper investigated two techniques of the *Bel Canto* tradition of operatic singing. Results indicate that male operatic singers do employ the "*aggiustamento"* or vowel modification, particularly in the upper limits of the pitch range. In this study soloists employed "*aggiustamento"* more so than chorus artists. However, all singers employed this technique in the upper pitch regions of the voice above 220Hz. Finally, evidence was presented that supported the notion of first formant tuning in the bass voice.

6. Acknowledgements

The authors would like to thank the State Opera of South Australia and Dr Frantz Clermont for their generous support in the production of this paper.

7. References

[1] Miller, R. (1996) The Structure of Singing: System and art in vocal technique. Schirmer – Thomson Learning, Belmont, USA.

[2] Appelman, D. (1967) The Science of Vocal Pedagogy. Reprint 1974. Bloomington, IN: Indiana University Press.

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

[4] Titze, I.R., Mapes, S.J., and Story, B.H., (1994) "Acoustics of the tenor high voice", J. Acoust. Soc. Am., 95(2), 1133-1142.

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

[6] Sunberg, J. (1975). Formant technique in a professional female singer". Acustica 32 89-96.

[7] Clermont, F. (2003) "Systemic Relationship between Bass Singers' Spoken and Sung Vowel Formants Spaces", 15th ICPhS, Barcelona.

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

[9] 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.

[10] Clermont, F. (1992) "Formant contour parameterization of vocalic sounds by temporally constrained spectral matching", Proc 4th Aust. Int. Conf. Speech Sci. & Tech., Brisbane. Pg 48-53.

[11] Sundberg, J. (1970) "Formant Structure and articulation of spoken and sung vowels", Folia Phoniatrica 22, Pg 28-48.

[12] Sundberg, J. (1974) "Articulatory interpretation of the "singing formant", J. Acoust. Soc. Am. 55, Pg 838-844.