How do they do it?
The difference between singing and speaking in female altos

Damien Hall*

1 Abstract

This paper reports the first set of results of an investigation into the phonetic correlates of female altos' singing (compared to the extensive research which has been done on males' singing). Based on the author's previous research (Hall 2003), predictions for this experiment were that when producing a 'singing' sound:

- singers would increase the amplitude of H2 to a level above the amplitude of H1
- at least the most experienced of the singers would produce a 'singer's formant'
- some formant tuning of F1 and possibly F2 would be observed: that is, singers would change the locations of F1 and F2 in order to amplify their sound. The degree to which they needed to do this would depend on f0 frequency.

(Two notes on terminology in this paper:

1) *H1* here means 'f0, the leftmost peak in the spectrum'; H2 refers to the peak above f0 in frequency.
2) The terms *f0* and *pitch* are interchangeable.)

Data were collected from nine singers in this experiment. Analysis in this paper concentrates on two of the nine: the singers who had the most and the least singing experience.

---

*I should like to thank my subjects, for being willing to let me analyse their voices; Rachel Smith, University of Cambridge Phonetics Laboratory, for an initial comment which inspired a good proportion of my analysis in work on this topic; Jonathan D Wright, University of Pennsylvania Linguistics Department, for much help, also with analysis; and William E Parberry, University of Pennsylvania Music Department, for passing on my request for subjects to the altos in two of his choirs.

The major finding is that female altos seem to use two techniques to amplify their singing sound:

- amplifying H2 so that the spectrum shows it to have a higher amplitude than H1 (this technique is referred to in this paper simply as 'amplifying H2')
- producing a singer's formant.

Which of these strategies is used depends on the individual singer, her amount of training / experience and the pitch-range in which she feels most comfortable.

2 Introduction

Beginning with Bartholomew (1934), much work has been done on the ways in which male singers produce their sound; there has been comparatively little on female singers, however, and work on the female alto voice is particularly lacking. One example of an experiment with alto subjects is the one reported in Sundberg and Skoog (1997), but the experimenters were not investigating the phenomena investigated here.

Hall (2003) found that the altos in that experiment used two techniques to make their 'singing' sound: amplification of H2 to a higher amplitude than H1, and production of a 'singer's formant'. The results suggested that singers used either one or the other of these techniques but not usually both: in general, the pattern seemed to be that as f0 rose, singers did not put energy into the spectrum at the frequency required to amplify H2, instead amplifying the higher-frequency area where F3, F4 and F5 would be found. Conversely, at lower f0 frequencies it was more common for singers to amplify their sound by raising H2 to a higher amplitude than by producing a 'singer's formant'.

2.1 Amplification of certain harmonics

Bartholomew (1934) first found amplification of the group of harmonics up to H5 by his baritone subject, and the same group of harmonics has been found to be amplified in a number of more recent experiments (Gramming 1991; Miller & Schutte 1990; Titze, Mapes & Story 1994). In the present experiment, amplification of H2 is common, though it is not necessarily the exclusive or preferred method of amplification of the overall sound.

2.2 The singer's formant

'Singer's formant' is the general name given to an area of high energy produced by singers and professional speakers between approximately 3kHz and approximately 4kHz (though the range can vary according to voice-type and individual voice, apparently as a mechanism for amplification and projection of the voice).

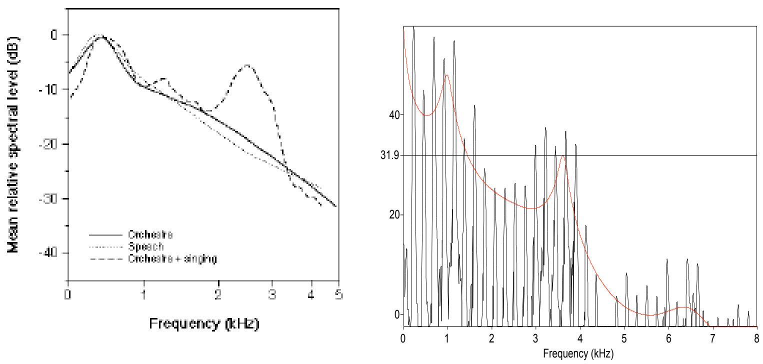

Figure 1: The singer's formant

Left: illustration in the dashed line ('orchestra and singing') (from National Center for Voice and Speech, nd)

Right: illustration from this experiment (FFT and LPC spectra) Singer A, vowel /a/, f0 ≈ 220Hz

LPC identifies a peak in the midst of a 'lump' shown in the FFT

2.3 Formant tuning

The term 'formant tuning' is used to refer to the deliberate approximation of some formant, usually F1 or F2, to a harmonic with which it would not usually align if left in the location where the formant naturally occurred. The movements in formants that are used to create a singer's formant, however, are of the same type, since a singer's formant is usually said to consist not only of an area of raised amplitude but of up to three formants which must actually have been moved closer together (ie away from their natural frequencies) in order to appear at the frequency necessary for the formation of a singer's formant. See Hall (2003): 'Sundberg (1973) remarks that it would be abnormal for the spectrum of a spoken vowel to contain five formants below 4000Hz, let alone below 3000Hz, but that such a density of formants is required (and attested) to explain the singer's formant.'

Little will be said of formant tuning in this paper, since it has not been possible so far to develop a reliable method of measuring significant differences in the position of a single formant; any impressions that formant tuning has happened are therefore based on visual evidence from spectrograms, not on statistical comparisons.

3 Method

3.1 Singers

Data were collected from nine female altos: the least trained had sung in a high school choir five years previously, but not sung since; the most trained had had seventeen years of professional training up to the present. This paper reports analysis of only the least- and the most-trained of the singers.

3.2 Equipment

Singers were recorded in the sound-proof booth at the University of Pennsylvania Phonetics Laboratory. A Shure SMS4 Condenser microphone was used, with a wind-shield; singers were asked to stand as close to it as was comfortable for them, which in most cases meant that their mouths were 15–20cm from the wind-shield and therefore 20-25cm from the microphone, though they moved their heads as they sang. Recording was direct to .wav format, using Adobe Audition on a PC, via an Apogee Electronics Mini-Me analogue-digital converter and a Mackie Micro Series 1202-VLZ 12-channel mic/line mixer, sampling at 44.1kHz. The only equipment in the recording booth with the singers was the microphone; everything else was outside, so there was no electronic background noise in the recordings. The experimenter sat outside the booth; communication between singers and experimenter was visual only.

Singers used a Farley's PocketTones Chromatic-C electronic pitch generator to give themselves pitches. Informal tests using Praat (Boersma and Weenink 1992-2005) showed that the pitch-generator consistently generated pitches within 0.01Hz–0.02Hz of their accepted values in the equal-tempered scale.

3.3 Materials / procedure

Singers were asked to produce the vowels in the English words *ah*, *air*, *gee*, *oh*, *ooh* (/a ɛ~e i o u/), first speaking them and then singing them.

- They were asked to speak the vowels using two intonations (demonstrated by the experimenter), quick rise-fall and quick fall-rise; each vowel-word was placed in sentence-initial position in a short carrier phrase, written on index-cards. This produced a total of ten spoken vowel-tokens for each singer, five with each intonation.

- They were then asked to sing each of the five vowels in two types of scale, *glissando* and *staccato*. Singers were again asked for upwards-downwards and downwards-upwards patterns; they were given index-cards containing all the scales needed.

  Glissando scales were sliding upwards (downwards) from a given note to the note a fifth above (below) it, and down (up) again: that is, from a given f0 to an f0 of 1.5 times (0.66 times) that value, and back again. *Staccato* scales followed the same pattern, except that they consisted of separated notes a semitone apart: for example, the notes sung in the upwards-downwards *staccato* scale beginning at 196Hz were 196, 208, 220, 233, 247, 262, 277, 294, 277, 262, 247, 233, 220, 208, 196Hz.

  The range tested was nominally G3–G4 (196Hz–392Hz), though some singers sang flat or sharp, so that the total range tested was slightly bigger.

  The scales elicited from each singer for each vowel were:
  - upwards-downwards scales starting at 196Hz, 208Hz, 220Hz, 233Hz, 247Hz and 262Hz, both *glissando* and *staccato* (2 styles x 5 vowels x 6 starting pitches = 60 scales)
  - downwards-upwards scales starting at 392Hz, 370Hz, 349Hz, 330Hz, 311Hz and 294Hz, both *glissando* and *staccato* (60 scales)

Singers were asked to speak the vowels before and after singing them. The difference between these two sets of spoken vowels, which can presumably be attributed to the voice being warmed up by singing, will be analysed in a follow-up paper.

In the sung part of the experiment, singers were given the choice of type of scale with which to begin, and, for each type of scale, the choice of whether to start at the top or the bottom of the range.

3.4 Analysis

Recordings were analysed in Praat version 4.2.34. The following analyses were used:

- wide-band spectrograms (0.017s)
- Linear Predictive Coding (Burg analysis, 16 poles) and Fast Fourier Transform spectra, generated from waveforms; a 50ms Gaussian window, giving an effective 100ms window, was used for spectra.

Singers' ability to keep pitch was not being tested in this experiment, but it was important that they should all sing at pitches which were roughly the same, so as to provide comparable results. During analysis, notes sung with an error of up to approximately 10Hz above or below the intended pitch were accepted as tokens of the intended pitch, though most singers' tuning was more accurate. Naturally, since rising f0 means that pitches are further apart in Hz, inaccuracy measured in Hz often tended to increase with increasing f0; however, in the vast majority of cases this did not affect the measurements that could be taken, since most singers' pitch was accurate.

For f0 frequencies in the middle of the range tested, singers produced several tokens in the course of the experiment. When this was the case, the following procedure was used to decide which of the tokens to analyse. It was assumed that the token with the highest amplitude would be the most 'singer-like' one, since the overall goal of the acoustic strategies investigated here is to increase the amplitude of a given sound (whether at certain frequencies in its spectrum or over the whole spectrum).

- The experimenter visually inspected the waveforms of the scales where the note in question was produced, to see whether the singer in question tended to produce greater amplitude as she sang higher or as she sang lower.
- If she produced greater amplitude at higher frequencies (which most did), the token of any given f0 frequency which appeared as the top note of an upwards-downwards *staccato* pattern was taken.
- If she produced greater amplitude at lower frequencies, the token of any given f0 frequency which appeared as the bottom note of a downwards-upwards *staccato* pattern was taken.

This strategy also meant that as many of the sampled notes as possible had been sung in the middle of the pattern, not at one end or the other. Both the decision to take samples from notes at the highest point of a pattern and the decision to take notes from the middle of a pattern where possible meant that the sampled notes tended to have higher amplitude, since many singers increased their amplitude towards the middle of a pattern and decreased it again towards the end. In any cases where the application of these two strategies produced conflicting results, the note with higher amplitude was preferred, whether or not it appeared in the middle of a pattern.

3.5 Measurements

Once the token from which measurements would be taken had been determined, measurements were taken as follows.

3.5.1 Amplitude of Harmonics

FFT spectra were created as described in §3.4 above. The difference in amplitude between the first two harmonics was calculated in Praat as the difference between the amplitude of a 20Hz bandwidth centred on f0 (ie starting 10Hz below f0 and ending 10Hz above it) and the amplitude of a 20Hz bandwidth centred on 2f0.

3.5.2 Singer's Formant / Formant Tuning

3.5.2.1 Location of Singer's Formant and Formant Tuning

The cursor was placed manually at the maximum amplitude in a given waveform which was also acceptably close to the intended pitch, and a Praat script was written to record the values of F1-5 at that position. Singer's formants were located visually from the FFT and LPC spectra created at that point. To measure formant tuning between different pitches in a single *glissando* scale, Praat scripts were written to automatically find the maximum and minimum pitches in that scale and record the values of F1-5 at those points.

3.5.2.2 Amplitude of Singer's Formants

LPC spectra were created and inspected visually for the presence of a singer's formant. Where a singer's formant was seen, two methods of measurement, both by hand, were tried:

- simple measurement of the amplitude of the LPC peak identified as the singer's formant
- measurement of the difference between the amplitude of the LPC peak identified as the singer's formant and the amplitude of the lowest point of the trough immediately to its left.

The method measuring the relative height of the singer's formant peak (the difference between the amplitude of the peak and the amplitude of the trough to its left) was chosen as more representative (see §4.2 below),

4 Results and Discussion

4.1 Amplification of H2

As Hall (2003) also found, singers often produced amplitude_{H2} considerably higher than amplitude_{H1} (they 'amplified H2') in sung vowels. In the following graphs, a positive value shows that the singer amplified H2 at the pitch indicated.

Straightforward comparison of these two singers' production for each vowel shows that the relationship between amplification of H2 and pitch is not as simple as it might appear. The results of Singer A (inexperienced) (Figures 2 and 3) might lead us to suppose that singers become better able to amplify H2 as pitch gets higher, but the results of the more experienced Singer I show that the relationship is more complicated. A trendline for Singer I's results (Figures 4 and 5) would still show that she amplified H2 to a greater extent overall as pitch rose, but the extent of her amplification is too variable to permit a strong conclusion relating (amplitude_{H2} - amplitude_{H1}) directly to pitch, since she crosses the *x* axis more than once for each vowel (ie there is not a unique transition between lack of ability / desire to amplify H2 and ability / desire to do so, as there seems to be for Singer A).

Interestingly, these results are contrary to those of Hall (2003), where the most experienced singers were found to amplify H2 *less* as pitch rose. The results from the current experiment indicate that we need to look further to define the relationship between amplification of H2 and pitch; if there is an answer, it may lie in the amount of experience a singer has, or simply in her preference of amplification strategy. For example, the more experienced singer analysed here, Singer I, may be relying more on a singer's formant to amplify her sound; this may not be for any articulatory reason but simply because she prefers to do so.

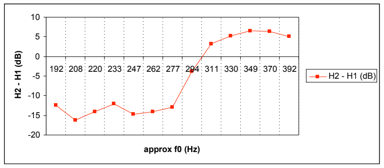

Figure 2: Singer A; vowel /a/ amplitude$_{H2}$ - amplitude$_{H1}$

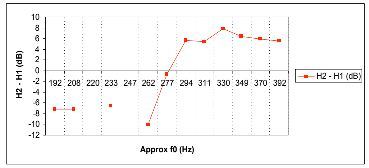

Figure 3: Singer A; vowel /o/ amplitude$_{H2}$ - amplitude$_{H1}$

(reliable tokens of intended 220Hz and intended 247Hz were missing for this singer for /o/).

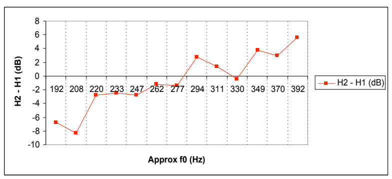

Figure 4: Singer I; vowel /a/ amplitude$_{H2}$ - amplitude$_{H1}$

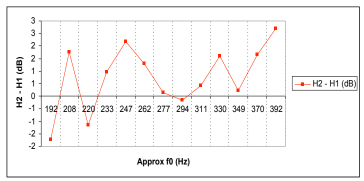

Figure 5: Singer I; vowel /o/ amplitude$_{H2}$ - amplitude$_{H1}$

4.2 Production of a singer's formant

Hall (2003) hypothesised that there might be a 'trade-off' between production of a singer's formant and amplification of H2; that production of a singer's formant might 'take over' as the method of producing a 'singing' sound when pitch was too high to permit amplification of H2. Figures 6-9 plot the difference between the amplitude of the singer's formant peak and the amplitude of the trough immediately to its left (since 'peaks' can be perceived as peaks only relative to the amplitude of the spectrum around them). They also show whether the singer did or did not amplify H2 in that token ('Y' or 'N').

The trendlines in Figures 6-9 allow us to compare production and level of the singer's formant with amplification of H2, to see whether production of a higher or lower singer's formant generally correlates with amplification of H2. R^2 values also showed that peak-to-trough measurement for singer's formants was a better measurement of ability to produce singer's formants than was the measurement of their absolute height. This is as expected because of the fact that peaks are seen as peaks only relative to the spectrum around them.

If a singer's formant is used as the preferred method of producing a 'singing' sound when f0 is not at a level which permits easy amplification of H2, conversely we would expect there to be no singer's formant at frequencies where the singers *can* amplify H2. Therefore, all the trendlines in Figures 6-9 would slope downwards from left to right, since it is at the right-hand end of these graphs, where f0 is higher, that these singers amplify H2. In fact, these measurements clearly do not permit any such easy 'trade-off' interpretation of the relationship between H2 amplification and production of a singer's formant, as hypothesised by Hall (2003). At this stage, with data from only two singers, only provisional conclusions are possible. However, we can observe that for both vowels, both the inexperienced Singer A (Figures 6 and 7) and the experienced Singer I (Figures 8 and 9) actually produce more of a singer's formant as pitch rises, and they also amplify H2 more.

From the point of view of accurate reflection of singing performance, the regression lines for singer's formant peaks relative to troughs still do not account for very much of the data. (Although the measurement of singer's formant peaks relative to troughs is generally an improvement on their measurement in absolute terms, the R^2 values indicate that the peak-to-trough measurement accounts for 25.6% of the data at best, among the samples analysed here.) A more accurate way of measuring the contributions of various singing techniques to producing a 'singing' tone is clearly necessary, but these methods are a good beginning.

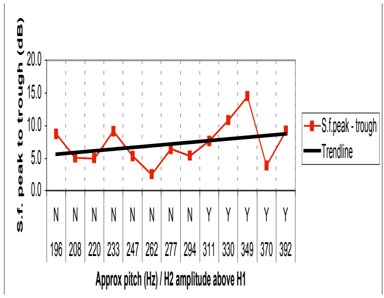

Figure 6: Singer A; vowel /a/
Singer's formant peak-to-trough distance

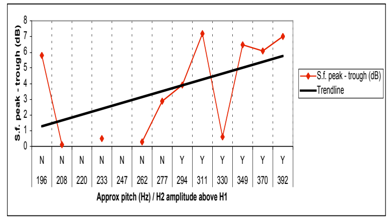

Figure 7: Singer A; vowel /o/
Singer's formant peak-to-trough distance

(Since these measurements were taken from the same set of tokens as the original measurements of (amplitude$_{H2}$ – amplitude$_{H1}$), again, reliable tokens of intended 220Hz and intended 247Hz were missing for this singer for /o/: cf Figure 3.)

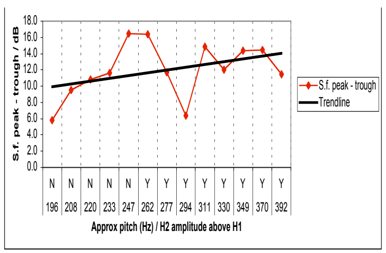

Figure 8: Singer I; vowel /a/
Singer's formant peak height

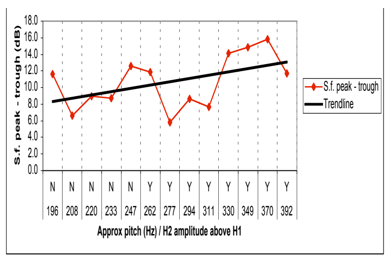

Figure 9: Singer I; vowel /o/
Singer's formant peak height

4.3 Formant tuning

This section will comment on tuning of F1 and F2, though, as has been said, the movements of F3-5 that seem to be necessary to create a singer's formant (moving F3-5 together and possibly also down in frequency as a whole) can also be seen as formant tuning. Movement of F3-5 could certainly be measured using whichever technique is used for F1 and F2.

Many of the *glissando* samples in this experiment exhibited formant tuning like the following:

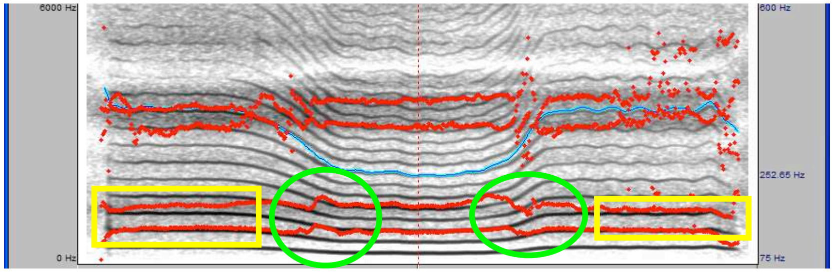

Figure 10: Singer A, vowel /a/, approx 392-262Hz

(In this figure, the thicker blue line following the (grey) harmonics is a pitch-track, measured on the scale at the right of the spectrogram (0-600Hz); the figure of 252.65Hz is the pitch at the point where the pitch-track crosses the cursor, in the middle of the figure, but that value is irrelevant here. The formants, in red, are measured on the scale at the left of the spectrogram (0-6000 Hz).)

The instances of formant tuning in Figure 10 illustrate two points.

1) The movement of formants from one harmonic to another, ringed in green, is an attempt by the singer to keep F1 and F2 at the same frequencies, when a drop in pitch is causing the harmonics generally to drop. Here, when the pitch drops, we see F1 moving from a position just above H2 to a position just above H3, and F2 moving from a position just above H3 first to just above H4 and then to just above H5; both formants move back to the harmonics they were originally closest to when the pitch rises back to the original level.

2) The sections boxed in yellow illustrate Titze's (2004) point that it may be more advantageous for a formant to be placed directly above a harmonic rather than at exactly the same frequency as a harmonic. In this figure, the left-hand box shows F1 positioned just above H2; the right-hand box shows F2 positioned just above H3.

Before substantive comments can be made on the issue of formant tuning, a systematic and objective way of measuring it for statistical significance will have to be found.

5 Conclusion

This preliminary study of two female alto singers with vastly different amounts of training and experience has shown that female altos do seem to use the two techniques previously found for the production of a 'singing' sound by male singers: the raising of H2 to a higher amplitude than H1 and the production of a singer's formant. It seems that the two techniques are often used together; the factors which influence the choice of technique and the extent of its use seem to be dependent on pitch, and may also be dependent on vocal training and simply on the individual singer's preference. Isolation of the factors which lead to the preference of one technique over another, if they are objective, will require further research on a larger body of data, to be presented in a follow-up paper. Likewise, the data for this study show that female altos, like male singers, do use formant tuning as a way to make a 'singing' sound, but precise investigation of the use and extent of formant tuning will come once a systematic and objective way of measuring it has been found.

References

Bartholomew, Wilbur J. 1934. A physical definition of good voice-quality in the male voice. *Journal of the Acoustical Society of America* 6(1):24-33.

Boersma, Paul, and Weenink, David. 1992-2005. *Praat* (version 4.2.34): software copyrighted by the authors. Downloaded from http://www.praat.org.

Gramming, Patricia. 1991. Vocal loudness and frequency capabilities of the voice. *Journal of Voice* 5(2):144-157.

Hall, Damien. 2003. *Formant tuning by female altos*. Ms., University of Cambridge.

Miller, DG, and Schutte, Harm K. 1990. Formant tuning in a professional baritone. *Journal of Voice* 4(3):231-237.

National Center for Voice and Speech. nd. Vocal Ring, or the Singer's Formant. Website accessed 6 November 2004: http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/singer.html

Sundberg, Johan. 1973. The source spectrum in professional singing. *Folia Phoniatrica* 25: 71-90.

——— and Skoog, Jörgen. 1997. Dependence of jaw opening on pitch and vowel in singers. *Journal of Voice* 11(3):301-306.

Titze, Ingo R. 2004. A theoretical study of F_0-F_1 interaction with application to resonant speaking and singing voice. *Journal of Voice* 18(3):292-8.

———, Mapes, S, and Story, B. 1994. Acoustics of the tenor high voice. *Journal of the Acoustical Society of America* 95(2):1133-1142.

Department of Linguistics
University of Pennsylvania
Philadelphia, PA 19104-6305
halldj@babel.ling.upenn.edu