VOWEL IDENTIFICATION IN SINGING AT HIGH PITCH

CW Thorpe[1] & CI Watson[2]
[1] School of Communication Sciences and Disorders, University of Sydney, and
[2] SHLRC, Macquarie University

ABSTRACT: We present a new analysis method that represents the vowel space directly by a factorial analysis of the harmonic amplitudes, without requiring explicit identification of formant frequencies. Analyses of vowels sung by male and female singers across their pitch ranges are performed with this method and also by LP formant extraction. The results indicate that even at high pitch, vowels are well separated with our new method, even though the LP analysis produces clusters of formants locked onto harmonic frequencies. This result suggests that vowel identity at high pitch may be conveyed largely by the magnitudes of individual harmonics, and that some of the observations of "vowel modification" and "convergence" in acoustic analyses of high pitch vowels may be artefacts of formant analysis.

INTRODUCTION

Vowel identity in speech is strongly correlated with the values of the first two or three formant frequencies (Fant 1960), which are formed by resonances in the vocal tract that produce peaks of energy in the vowel spectrum. In singing however, the situation is complicated at high pitch because the frequency separation between voice partials can be comparable to the formant frequencies. For instance, Concert A has harmonics separated by 440Hz. which means that there are only two partials across the usual range of the first formant frequency F1 (300 to 800 Hz in speech). In order to maintain a desirable vowel quality, singers tend to modify their vocal tract configuration to improve the match between voice source spectrum and the vocal tract resonances. Several studies have examined this difference in formant structure between the speaking and singing voice (e.g. Sundberg, 1970), and the effect of high pitch on the vowel characteristics (Di Carlo & Germaine, 1985). Sundberg (1970) found that in singing, the formant frequencies of the front vowels were lowered, due to the effect of a lowered larynx during singing. At higher pitch, F1 and F2 for the high vowels appeared to increase in conjunction with the frequencies of the first two partials, whereas F3 decreased (Sundberg, 1975), this effect being explained by an increase in jaw opening as pitch increased. Accurate analysis of this phenomenon is however difficult because one cannot observe the vocal tract resonances themselves, but only the voice output which is the convolution between the vocal tract resonance and the periodic glottal excitation signal. The spectrum of the voice signal therefore contains energy only at multiples of the fundamental frequency, implying that the vocal tract resonance is sampled at these points only. Unfortunately, when the fundamental frequency is high (i.e. comparable or higher than the centre frequencies and bandwidths of the formant resonances) there is insufficient information to reconstruct the shape of the resonance spectrum. There is some evidence that perception of vowels at high pitch is modified to take into account this fact of missing information (de Cheveigné & Kawahara 1999).

Formant analysis using linear prediction (LP) is essentially a process of modelling the voice spectrum with a certain small number of resonance peaks. However, when the number of harmonics is already small, LP analysis often simply identifies one or more of these as narrow bandwidth "resonances". For high vowels in particular, this begins to occur at relatively low pitch, giving the effect that the apparent value of the first formant is locked to the fundamental frequency. The question then arises as to how to separate this analysis artefact from the phenomenon of "vowel modification". In this study, we investigate changes in vowel shape as pitch increases by direct analysis of the harmonic magnitudes.

METHODS

Sung Italian vowels were recorded from two classically trained professional singers, a soprano and a counter tenor. Sustained notes (isolated vowels) and sustained vowels in a VdV context were sung over a pitch range from A3 (220Hz) to C#5 (554Hz) by the male singer and E3 (165Hz) to G#5 (830Hz) by the female singer in a sound treated audiometry booth. Sounds were recorded on DAT tape and digitally transferred to computer at a sampling rate of 44.1kHz. The vowels were labelled using the Emu speech

data analysis system (Harrington & Cassidy, 1999). The acoustic vowel onset was marked at the onset of voicing, as shown by strong vertical striations in the spectrogram, and the offset was marked at the cessation of periodicity or a substantial decrease in waveform amplitude (Watson & Harrington, 1999).

Because each subject sung in a different key, the nearest applicable notes were combined. For the results presented here, we show plots with vowels sung at G#3 and A3 (F0 = 207 and 220Hz respectively), G#4 and A4 (F0 = 415 and 440Hz respectively) and at G#5 (F0 = 830Hz). The vowels from both the VdV and the isolated contexts were grouped for this analysis. This gave us a total of 103 tokens in each of octave 3 and 4, and 45 tokens in octave 5.

Formants were computed using 12th order autocorrelation LPC in ESPS Waves (Entropic), with 10ms segments spaced at 10ms down-sampled to 10kHz. Tracking errors in F1 and F2, where portions of the formant track were labelled by Waves as the formant either above or below the dominant formant through the vowel, were manually corrected in Emu. However, formant tracks that were at phonetically unusual frequencies were not changed if they were consistent throughout the vowel. The frequencies of the first two formants were obtained from the mid-point of each vowel.

Power spectra were computed from the voice signal by performing fast Fourier transforms (FFT) on 23ms segments (Blackman-Harris window, 8192 point zero padded FFT) spaced 10ms throughout the duration of the vowel. For each spectral slice, the fundamental frequency (F0) was automatically identified by maximising the sum of the harmonic magnitudes. After visual inspection of the F0 track, the power of each of the first 10 harmonics was measured and averaged across the duration of the vowel.

Principal component analysis (PCA) was performed on the sets of harmonics at each note using the Statview (SAS Institute) analysis package. Factors with eigenvalues greater than one were retained and the varimax rotation applied. Bivariate scattergrams of the individual data points were then plotted on the orthogonal axes with each point labelled with its vowel identity.

RESULTS

Figure 1 shows the formant charts for the lowest octave. Vowel clusters are well separated and in the expected positions for Italian vowels. At the second octave (Figure 2), some clustering of the formant values is occurring at the frequencies of the harmonics, with F2 for subject 2 tending to cluster at 880 and 1760Hz, and F1 for subject 3 appearing at around 400 and 800Hz. At the highest octave, with F0=830Hz, both formants are tightly clustered around the fundamental and first harmonic.

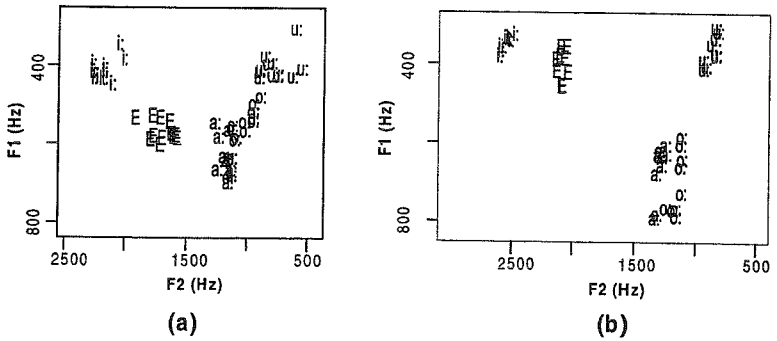

Figure 1. Formant plots for the vowels sung in the lowest octave. (a) Subject 2, male (F0=220Hz - A3). (b) Subject 3, female (F0=207Hz - G#3).

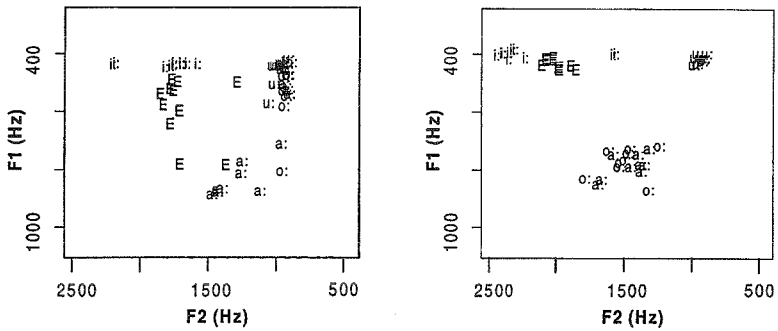

Figure 2. Formant plots at the second octave for (a) male singer at A4 (440Hz) and (b) female singer at G#4 (415Hz).

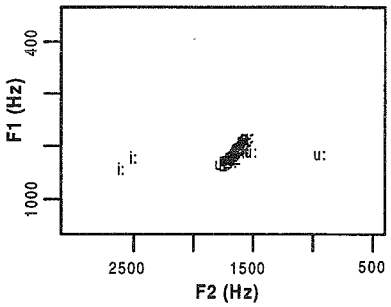

Figure 3. Formant plot for sung vowels at a pitch of G#5 (F0 = 830Hz).

The PCA at the lowest octave (G#3/A3) resulted in two factors that accounted for 82% of the variance. At the next octave (G#4/A4), three factors with eigenvalues greater than one accounted for 86% of the variance, while at the highest octave (G#5), three factors account for 75% of the variance (note only a single subject in this octave). The factor weightings for each of the 10 harmonic components are detailed in Tables 1-3 respectively for the three octaves, and vowel charts showing the position of the sung vowels with respect to the orthogonal factors are shown in Figures 4 through 6.

At the lowest octave Factor 1 represents the harmonic amplitudes from 200 to 1200Hz whereas Factor 2 represents harmonics from 1400Hz to 2000Hz (note these frequencies are rounded for convenience). As shown in Figure 4 there is a clear separation of vowel clusters, with a very similar pattern to the F1-F2 vowel chart. At the next octave, Factor 1 represents the contribution of harmonics between 2800 and 4000Hz, Factor 2 of harmonics between 1600 and 2400Hz, and Factor 3 the low frequency harmonics between 400 and 1200Hz. The first factor provided a distinction between the subjects, so Figure 5 shows the vowels against Factors 2 and 3. Again there is a clear separation between the vowels, although the two subjects occupy somewhat different spaces. The vowel clusters for each subject are still however arranged according to the usual vowel chart pattern. At the highest octave (harmonic spacing 830Hz), there is significant overlap of the vowel clusters, although still with a clear front/back and some high/low distinction. At this pitch, Factor 1 represents harmonics between 800 and 3200 Hz, Factor 2 harmonics between 4800 and 8000 Hz, and Factor 3 several miscellaneous minor weightings.

Harmonic	(approx Hz)	Factor 1	Factor 2
A0	(200Hz)	**-.8**	0
A1	(400Hz)	**-.8**	-.2
A2	(600Hz)	**.9**	.1
A3	(800Hz)	**.9**	-.3
A4	(1000Hz)	**.9**	0
A5	(1200Hz)	**.9**	.4
A6	(1400Hz)	.5	**.8**
A7	(1600Hz)	.2	**.9**
A8	(1800Hz)	-.1	**1.**
A9	(2000Hz)	-.2	**.9**

Table 1. Orthogonal factor scores for the PCA at octave 3 (G#3/A3).

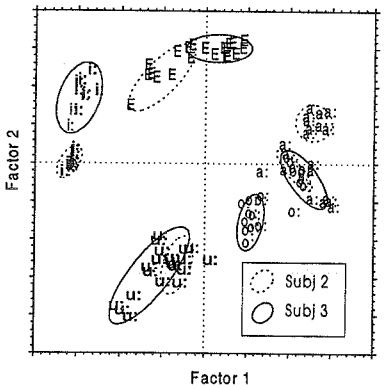

Figure 4. Vowel tokens at the lowest octave (G#3 /A3) plotted against the two principal components obtained from the 10 harmonic amplitudes.

Factor interpretation

From the results of the PCAs conducted on the three sets of vowels, it is apparent that the component distinguishing front and back vowels (Factor 2 in octaves 1 & 2) represents harmonic energy at frequencies between 1400 and 2500Hz. These are vowels with a high F2 within this frequency region. At the highest octave, the distinguishing component is Factor 1, which again has positive weights on components between 1600 and 3200Hz, but with a compensating negative weight on the fundamental at 800Hz. Because there is only a single partial below 1400Hz, this negative weight reflects the fact that for the back vowels both F1 and F2 are low. The main distinction in this PCA vowel chart (Figure 6) is therefore between the front and back vowels.

High and low vowels in the first two octaves are distinguished by a component with negative weights on low frequency harmonics (below 400Hz) and positive weight on harmonics between 800 and 1200Hz. This provides a dimension that approximates to the range of F1. At the highest octave however, the distinction is provided, albeit in much reduced form, by a combination of higher harmonics.

Harmonic	(approx Hz)	Factor 1	Factor 2	Factor 3
A0	(400Hz)	0	.2	**-.8**
A1	(800Hz)	-.2	**-.5**	**.8**
A2	(1200Hz)	-.2	.3	**.8**
A3	(1600Hz)	.5	**.8**	.2
A4	(2000Hz)	.3	**.9**	-.2
A5	(2400Hz)	-.3	**.9**	-.2
A6	(2800Hz)	**.8**	.3	-.1
A7	(3200Hz)	**.9**	0	.2
A8	(3600Hz)	**1.**	-.1	-.1
A9	(4000Hz)	**.9**	0	-.3

Table 2. Orthogonal factor scores for the PCA at octave 4 (G#4/A4).

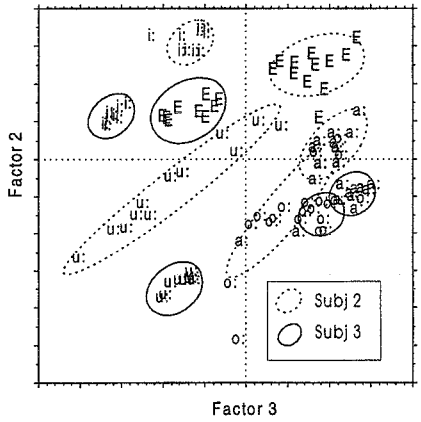

Figure 5. Principal component vowel chart of vowels sung at the second octave (G#4 / A4, F0 = 415/440Hz)

Harmonic	(approx Hz)	Factor 1	Factor 2	Factor 3
A0	(800Hz)	**-.9**	.2	-.2
A1	(1600Hz)	**.7**	-.3	.4
A2	(2400Hz)	**.7**	.3	-.5
A3	(3200Hz)	**.8**	.1	.1
A4	(4000Hz)	.5	-.1	0
A5	(4800Hz)	0	**.8**	.2
A6	(5600Hz)	.1	.4	**.8**
A7	(6400Hz)	.6	**.6**	-.3
A8	(7200Hz)	0	**.9**	0
A9	(8000Hz)	-.3	**.9**	.1

Table 3. Orthogonal factor scores for the PCA at octave 5 (G#5).

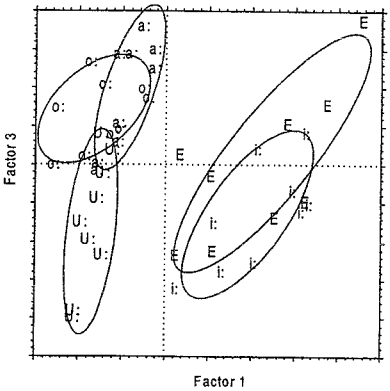

Figure 6. Principal component vowel chart for vowels sung by a female singer at a pitch of G#5 (F0=830Hz).

CONCLUSIONS

It is shown that sung vowels can be separated by an analysis of their harmonic magnitudes, even at high pitch where formant analysis leads to clustering of the formants at multiples of the fundamental frequency. The vowel charts so obtained can be related to traditional vowel production patterns, with the factor components representing concentrations of harmonic energy in different frequency bands corresponding to the expected locations of the formant resonances.

With regard to the question of vowel modification in singing, the results here indicate that there may be some merging of the high and low vowels, at least as far as the spectral shape is concerned, but that the vowel space is not reduced as much as is implied by the harmonic induced quantising of the formant frequencies as occurs in traditional formant analysis.

This method may also have implications for the analysis of high pitch speech, for instance in children or women's voices where F0 can be as high as 400Hz. Although accurate determination of the pitch is necessary, the harmonic analysis method does not require hand correction. This has potential applications in real-time situations, for instance for visual feedback vowel training.

REFERENCES

de Cheveigne, A. & Kawahara, H. (1999). "Missing-data model of vowel identification", *J. Acoust. Soc. Am.* **105**, 3497-3508.

Fant, G. (1960). *Acoustic Theory of Speech Production.* The Hague, Mouton.

Harrington, J. & Cassidy, S. (1999) *Techniques in Speech Acoustics.* Dordrecht, Kluwer.

Scotto Di Carlo, N. & Germain, A. (1985). "A perceptual study of the influence of pitch on the intelligibility of sung vowels", *Phonetica* **42**, 188-197.

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

Sundberg, J. (1975). "Formant technique in a professional female singer", *Acustica* **32**, 89-96.

Watson, C.I. & Harrington, J. (1999) "Acoustic evidence for dynamic formant trajectories in Australian English vowels", *J. Acoust. Soc. Am.* **106**, 458-468.