# SPECTRAL MEASUREMENT OF VOICE QUALITY IN OPERA SINGERS : THE CASE OF GRUBEROVA

*Tom Johnstone and Klaus Scherer*
*University of Geneva*

## ABSTRACT

Excerpts from different recordings of the cadenza in Ardi gli incensi from Donizetti's opera Lucia di Lammermoor as sung by Gruberova are acoustically analyzed to determine the nature of higher frequency energy and higher formant structure (in particular the presence or absence of a singer's formant) particular to this singer. In light of the results, the role played by these acoustic features in the expression of emotion in opera singing is discussed.

## INTRODUCTION

In a recent study, Siegwart and Scherer (1995) acoustically analyzed two excerpts from the cadenza in Ardi gli incensi from Donizetti's opera Lucia di Lammermoor as sung by five famous sopranos (del Monte, Callas, Scotto, Sutherland, Gruberova). The acoustic parameters measured were correlated with preference and emotional expression judgments, based on pairwise comparisons, made by a group of experienced listener-judges. In addition to showing major differences in the voice quality of the five dive studied, the acoustic parameters suggested which vocal cues affect listener judgments. Two component scores, based on a dimensional analysis of the acoustic parameters, predicted 84% of the variance in the preference ratings.

The results showed that Gruberova's rendering of the cadenza was generally preferred and thought to express more "tender passion" and "sadness" than the other singers by the judges in this study. Acoustically, Gruberova's voice was markedly different from all the other singers, showing lower energy in a predicted singer's formant band and stronger high frequency energy components in the spectrum. Her voice was like that of Sutherland and Callas in being characterized by high energy in the F0 band and little variation in lower frequency peaks.

Reviewers of the above study raised the possibility that the strong higher frequency energy measured for Gruberova (which might have affected the judges ratings) could be due to the sound engineers' selective boosting of specific frequency bands. To study this possibility, we recorded the two excerpts of the cadenza live in Gruberova's dressing room before a performance of Lucia at the Zurich opera, having obtained the artist's cooperation for this study. In this paper we compare the acoustic analyses for this recording with several professional recordings - a new CD recording and a radio broadcast from a concert hall, in addition to the cassette recording used in the earlier study. The recording of Sutherland (1971 CD) from the original study is also included for comparative purposes.

This paper reports the acoustic results and reviews the role of the higher frequency spectral energy bands and the singer's formant for the expression of emotion in speech and singing.

## METHOD

Gruberova's rendering of the two lines of the Lucia cadenza studied in this research were recorded in her dressing room using a Sony TCD-D3 DAT recorder. The prerecorded samples consisted of a 1984 EMI cassette, a 1992 Teldec CD, and a recent live radio recording. The recordings were digitized using a Kay CSL 4300B speech station at 20kHz sampling rate. An optimal recording level was chosen for each sound recording. This did not affect the subsequent analyses, as the spectral measurements of intensity were all normalized with respect to the total intensity of each recording.

## RESULTS

Analysis of the digitized recordings paralleled that used in the original study [1]. A 128 point long term average spectrum was calculated for the full duration of each of the digitized recordings. This spectrum was used to calculate the relative amount of energy present in the two frequency bands measured in the original study (Table 1, rows 1 and 2). With the exception of the CD recording, the energy in the high frequency band (3500 to 10000 Hz.) is higher in the Gruberova recordings than in the recording of Sutherland (Figure 1).

The CSL LPC formant tracking program was then applied to each recording, yielding means and standard deviations of the frequencies of tracked formants (Table 1, rows 4 to 7). The figures for the fourth tracked formant are included for purposes of comparison with the data from the original study. The figures pertaining to this formant should be regarded with some caution however, as the formant was not consistently identified by the tracking algorithm and the number of valid samples varied substantially between the different recordings.

The formant figures also show general agreement with the original study. Specifically, the recordings of Gruberova consistently show a third formant located at a higher frequency than that of Sutherland. The recordings of Gruberova also have a higher standard deviation of the fourth formant than the Sutherland recording, as in [1]. Contrary to the previous study however, there was no significant difference between the two singers in the mean positions of the fourth formant.

An examination of the relationship between the different measured parameters can help determine the nature of the measured high frequency energy. Correlations were calculated between the two formant frequencies and the energy in the first two frequency bands. It was found that the frequency of the third peak correlates strongly with the amount of energy in the spectrum above 3500 Hz (Pearson's $r = 0.97$). Examination of the long term average spectrum for the five recordings indicates that the amplitude decreases sharply above 4500 Hz., indicating that energy in this region does not contribute substantially to the high frequency band measured. Thus it would seem that the large amount of high frequency energy in the Gruberova recordings is due to the higher frequency position of the third spectral peak. This was confirmed by measuring the energy in the frequency range from 3500 to 4500 Hz. (Table 1, row 3). All the Gruberova recordings were characterized by more energy in this band than the recording of Sutherland. Importantly, Sutherland's (and the other *Dive* studied in [1]) third formant is located under the 3500 Hz. cutoff for the two measured high frequency bands.

An examination of spectrograms can be used to better understand the nature of the formant structure in the different recordings. Figure 2 shows the spectrograms for the live recording of Gruberova and the recording of Sutherland (only one spectrogram of Gruberova is displayed here, although spectrograms of the other Gruberova recordings were very similar, showing concentrations of energy in the same regions). The spectrograms reveal quite a different spectral energy distribution for Gruberova as compared to Sutherland. Gruberova shows a concentration of energy in two closely spaced bands between 2900 and 4100 Hz., with relatively low energy in the 1500 to 2500 Hz. range. In comparison, Sutherland shows a more constant spectral slope, with three bands or reducing energy located between 1500 and 3900 Hz. It is also apparent that the automatic formant tracking program was not able to distinguish between the third and fourth formants of both singers, thus compounding the two into one measured formant track (the third formant as given in Table 1, rows 4 and 5).

## DISCUSSION

The original purpose of this study was to determine whether or not the presence of more high frequency energy and higher frequency peaks in recordings of Gruberova was due to recording artifacts or manipulation by sound engineers. By analyzing three new recordings, including one taken directly in the singer's dressing room, it has been shown that the high frequency energy is indeed a characteristic of Gruberova's singing itself. More specifically, long term spectra of all Gruberova recordings displayed a high energy region between 2900 and 4100 Hz. This region appears due to the clustering of the third and fourth formants. In contrast, the recording of Sutherland lacks such a high energy region and the formants appear at lower frequencies.

An explanation of the high energy region in the recordings of Gruberova might be the presence of a singer's for-

*Table 1. Acoustic analysis results for the five recordings . Rows 1-3 are given in dB, normalized with respect to overall intensity. Rows 3-7 are given in Hertz.*

| | Gruberova | | | | Sutherland |
|---|---|---|---|---|---|
| | Cassette | Live | CD | Radio | CD |
| Singers Formant | -36 | -46 | -32 | -36 | -36 |
| 3500-10000 Hz. | -28 | -32 | -36 | -27 | -36 |
| 3500-4500 Hz. | -29 | -32 | -32 | -27 | -38 |
| Peak 3 Mean | 3779 | 3618 | 3546 | 3797 | 3443 |
| Peak 3 Std. Dev. | 236 | 399 | 321 | 229 | 355 |
| Peak 4 Mean | 4982 | 4826 | 5427 | 4934 | 5320 |
| Peak 4 Std. Dev. | 1640 | 1253 | 1981 | 1427 | 403 |

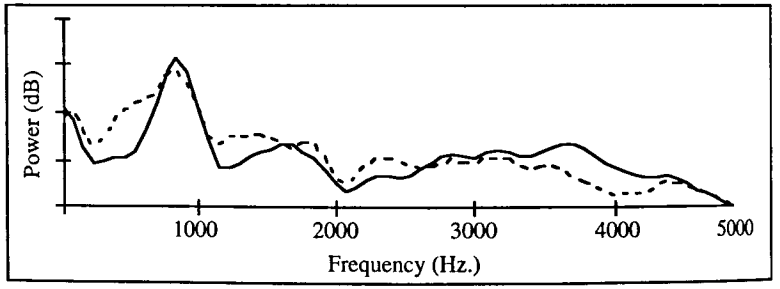


*Figure 1. Average spectra for Sutherland (broken line) and live recording of Gruberova (solid line). The spectra are normailized to overall intensity.*

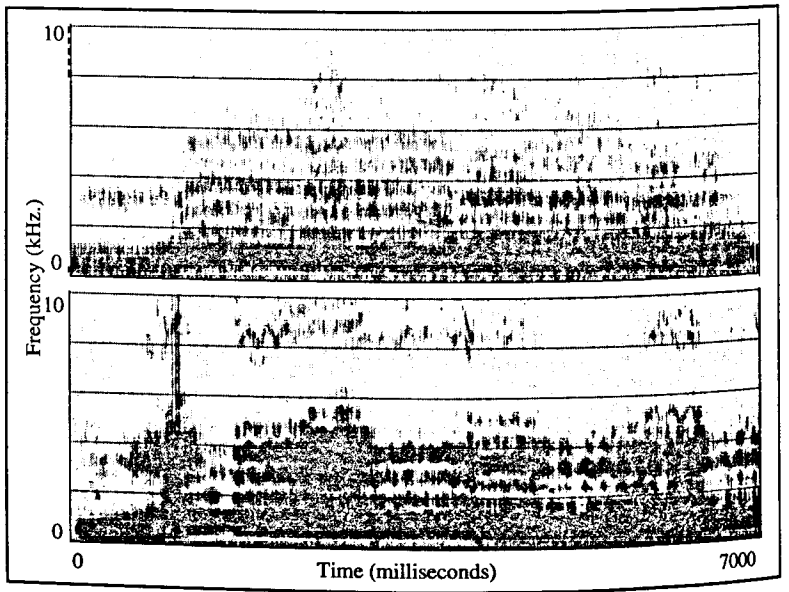


*Figure 2. Normalized spectrograms of Sutherland (top) and Gruberova (bottom).*



mant centred at about 3600 Hz. As discussed in [2], the singer's formant is not a single formant as such, but rather a clustering of formants around a predicted frequency of about 3000 Hz (in sopranos). When clustered sufficiently closely, individual formants tend to reinforce each other, leading to a spectral region with increased overall resonance. In the case of soprano singing, the partials are spaced widely apart, which makes the exact positions of the formants relative to the partials crucial.

Whilst most sopranos may be able to vary the formants to follow the positions of the harmonics, the way in which this is done may vary between singers. Thus some singers might raise the fourth formant in order to make it coincide with a harmonic, thus separating it from the lower formants, which typically might drop ([2], pp. 125-129). Such a separation of third and fourth formants, which would prevent the development of a singer's formant, would seem to be the case with Sutherland. In the recordings of Gruberova, however, the fourth formant drops along with the third formant, thus maintaining a close distance and allowing the formants to reinforce. The presence of a singer's formant will not necessarily ensure high energy in that region of the spectrum; the spectral drop-off of the harmonics must also be sufficiently gradual.

The question of whether sopranos possess a singers formant has been discussed recently by Berndtsson and Sundberg [3]. Berndtsson and Sundberg compared the classification by trained judges of synthesized soprano voices for various manipulated singer's formant positions. Also included in the study was one recording resynthesized using the formant positions from a professional soprano. The study found that perceived quality of the synthesized voices increased as the centre frequency of the singer's formant increased. The recording resynthesized from the professional soprano's formant positions was, however, judged as natural as the best of the synthesized recordings, despite its lack of a singer's formant.

The strong correlation found in [1] between the proportion of energy above 3500 Hz. and judgments of emotional expressivity may well be due to the presence of a singer's formant at about 3600Hz. This would fit well with the results using synthesized recordings in [3]. As that study only examined singer's formant positions up to 3500 Hertz, it was unclear whether even higher positions for the singer's formant might be judged even better. This study indicates that perceived quality and expressivity might continue to increase with an even higher singer's formant. The finding in [3] that the resynthesized recording with no singer's formant was judged to be as natural as the synthesized recordings might have been due to the more natural formant spacing, rather than the lack of a singer's formant per se. The relative positions of the formants in relation to the harmonic structure might be crucial to perceived quality, and thus the synthesized recordings using formant spacing taken from baritones might have suffered from their somewhat arbitrary relative formant positions. As admitted by the authors of that study, none of the recordings in their study were judged as being particularly natural.

Many of the conclusions drawn here concerning higher frequency spectral regions and the formant structure of sopranos remains speculative. In particular, the temporal changes to these features have not been examined. It is clear that much further empirical research is required in order to better understand the processes involved in emotional expression in singing.

## ACKNOWLEDGMENT

The authors wish to express their gratitude to Editha Gruberova for her cooperation in this study.

## REFERENCES

[1] Siegwart, H. and Scherer, K. R. (1995), "Acoustic concomitants of emotional expression in operatic singing: The case of Lucia in Ardi gli incensi," *Journal of Voice.*

[2] Sundberg, J. (1987), "The science of the singing voice," Dekalb, Il: Northern Illinois University Press.

[3] Berndtsson, G. and Sundberg, J. (1995), "Perceptual significance of the centre frequency of singer's formant," *STL-QPSR*, 4/1994, KTH, Stockholm, pp. 95-105.