Acoustic Concomitants of Emotional Expression in Operatic Singing: The Case of Lucia in Ardi gli incensi

Siegwart-Zesiger, Hervine Marie; Scherer, Klaus R.

Journal of Voice
Vol. 9, No. 3, pp. 249–260

Acoustic Concomitants of Emotional Expression in Operatic Singing: The Case of Lucia in *Ardi gli incensi*

Hervine Siegwart and Klaus R. Scherer

Department of Psychology, University of Geneva, Geneva, Switzerland

Summary: Two excerpts from the cadenza in *Ardi gli incensi* from Donizetti's opera *Lucia di Lammermoor* were acoustically analyzed for five recorded versions of the cadenza by Toti dal Monte, Maria Callas, Renata Scotto, Joan Sutherland, and Edita Gruberova. These acoustic parameters of the singing voices 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 factorial–dimensional analysis of the acoustic parameters, predicted 84% of the variance in the preference ratings. **Key Words:** Singing voice—Emotional expression—Acoustic analysis—Opera performances.

Ever since opera was born at the time of the Florence camerata in the 16th century, it has been suggested that the major task of this type of music is to express emotion. In this tradition, Kivy (1), in a scholarly analysis of the historical development of opera, has attempted to show how the philosophical–psychological notions about the nature of emotion that are held in a particular period influence the rendering or the representation of the emotions in operatic form. His argument is that baroque opera, particularly Händel, uses the *da capo* form to run through complete cycles of clearly defined and delimited emotions, in line with the Cartesian notion of a limited number of "passions of the soul," whereas Mozart's operas, using the *sonata* form, are imbued with the spirit of English associationist ideas about emotional fluidity and changeability. How exactly emotion is composed into opera is a fascinating question for musicologists and for anyone interested in the history of ideas and of the arts.

However, there is also an interesting issue for performance research in the psychology of music and voice science. The emotion composed into *recitativos, arias,* and ensembles by the composer is then portrayed by the singer. Presumably, a highly complex mixture of factors, (including the personality of the singer, the interpretation of the piece by the director, and the emotional climate in the opera house on a particular evening) determine the way the character and his or her emotions at particular points in the action will come to life (and to sound) through the interpretation of the singer.

In the light of the importance of emotional feeling and emotional communication on the operative stage, it is surprising to find so little pertinent research on this topic in the psychology of music. The lack of evidence becomes more understandable if one considers that, even for theater where there has been concern with this issue for centuries [see, for example, the analyses of affect display in the theater provided by Greek and Roman philosophers, or Engel (2)], there is little consensus and even less research evidence on the issue of how emotions are to be expressed on the stage. This is true for the

Accepted February 4, 1994.
Address correspondence and reprint requests to Dr. K. Scherer at Department of Psychology, University of Geneva, 9, route de Drize, CH-1227 Carouge-Geneva, Switzerland.
A preliminary version of this paper was read at the 2e Convegno Europeo di Analisi Musicale, Trento (Italy), October 24–27, 1991 (Proceedings published by Universita degli Studi di Trento, Dalmonte R, Baroni M, eds. Trento, Italy, 1992).

basic attitude or technique required (e.g., Stanislavski versus classical training) or the vocal, facial, or gestural patterns to be employed for the portrayal of specific emotions (3–5).

The issues are even more complicated in opera, because the expression of naturalistic emotions is constrained by historical and stylistic factors (e.g., the structure of the arias and musical symbolism in Baroque operas, Belcanto technique, *Sprechgesang*, etc.) as well as the technical requirements for the execution of the melody. Although the problem of emotional expression in opera singing is complex, it is not beyond scientific study. Although there have been few studies of emotional singing (but see ref. 6), there has been a fair amount of work on emotional expression in speech (see review, refs. 7–9), and the approach taken in this area of research could provide a model for similar work on emotional expression in singing.

As in speech research, three major questions can be asked in studies on the vocal expression of emotion in music:

1. How does an emotional state, with all the accompanying physiological effects of respiration, phonation, and articulation, manifest itself by systematic changes in the acoustic parameters of the voice? Or, in the case of a singer portraying the affective state of a character, how does he or she change the acoustic features of the voice to convincingly render the particular emotional state? In speech research, either recorded samples of speech occurring during real emotional upheaval or actor portrayals for different types of emotions are obtained, and systematic analyses of the respective acoustic features are conducted to determine the nature of the acoustic effects. The results of this type of research demonstrate that there are specific patterns of acoustic cues that characterize particular emotional states (3,8).
2. Can listeners correctly infer the nature of the underlying emotion (real or portrayed), based only on the acoustic cues in the singing voice? In speech research, to answer this question, emotional voice samples are presented to listener-judges who are then asked to choose the emotion expressed among a set of emotion categories. This type of research has shown an accuracy of recognition (~50–60%) that is approximately five times higher than what one would expect by chance (10,11).
3. What are the acoustic cues that listeners use to infer the nature of the expressed emotion from the singing voice? To elucidate this question, speech researchers have developed elaborate research designs using procedures of partial masking or filtering of specific cues (12), or acoustically measured cues have been correlated with listener judgments that have been obtained for the same voice samples (4,13). In more recent studies, (rule) synthesis and resynthesis (copy synthesis) methods have been used to study the issue, yielding a number of important findings on cue use in emotion inference from the voice (14).

Clearly, similar types of studies are of interest in exploring operatic performance with respect to emotional expression of the artist. How does a singer vocally "encode" the emotional state of a character at a particular point in the drama? What does the audience infer from this set of acoustic features with respect to the emotional states that are communicated? Are there some acoustic features that are particularly important in determining the audience judgment and preference?

This article attempts to contribute to the development of research on emotion expression in operatic singing by measuring acoustic features in existing recordings of operatic performance and linking them to audience evaluation. In a case study, meant to explore the possibilities for empirical research in this domain, we used parts of the cadenza after the cadenza *Ardi gli incensi* in the "mad scene" from *Lucia di Lammermoor* by Donizetti.[1]

Using several recordings of performances of the cadenza by different sopranos, we attempt to answer the following questions:

1. How do the artists studied differ in their vocal portrayal of the emotional state of Lucia in the "mad scene"? Although the emotion is not unambiguously defined by either the author of the libretto or the composer and because the intentions of the director and the artists are not known in the case of recorded material, we cannot link the acoustic features found to a particular emotional state. However, differ-

[1] This had been originally suggested by the organizers of 2e Convegno Europeo di Analisi Musicale (see last footnote on first page of this article) as a common task for the participants of a symposium on different approaches to music analysis.

ences in the vocal rendering of the cadenza might yield cues as to the artist's interpretation.

2. Which of the vocal portrayals of Lucia's emotional state are preferred by experienced opera listeners and what underlying emotions do they believe to be expressed?
3. What are the acoustic features in the singers' voices that seem to have the greatest impact on the listener judgments; in other words, what cues seem to be used to infer an emotion or to prefer a specific voice?

METHOD

Two short excerpts from the cadenza were chosen for systematic judgment studies and acoustic analysis (see the respective parts of the score in Fig. 1). The choice of the two excerpts was motivated in part by considerations of practicability, such as little instrumental accompaniment, and in part by the concern to have two different melodic structures represented.

We then digitized these excerpts from five recorded versions of the cadenza, by Toti dal Monte (1926, MC), Maria Callas (1955, CD), Renata Scotto (1967, CD), Joan Sutherland (1971, CD), and Edita Gruberova (1984, MC). During most of the excerpts, the flute accompaniment (Fig. 1) was audible. Digitalization was performed by using an OROS A/D converter with a sampling frequency of 20 kHz on a COMPAQ 386 PC. Acoustic analysis was performed with the help of the SPECTRO analysis package (developed in our laboratory). Specifically, we computed the energy envelope, the sequence of spectral peaks (using a peak-picking routine equivalent to the one used in the ILS analysis package), and the average 128-point spectrum in the 78–10,000-Hz range, for each of the two samples selected.

It should be noted that because the acoustic analyses were performed on the basis of the line output signal of the recording, the parameters measured, particularly the energy level and distribution, reflect the technical settings of the sound recording and mastering as well as the performance of the

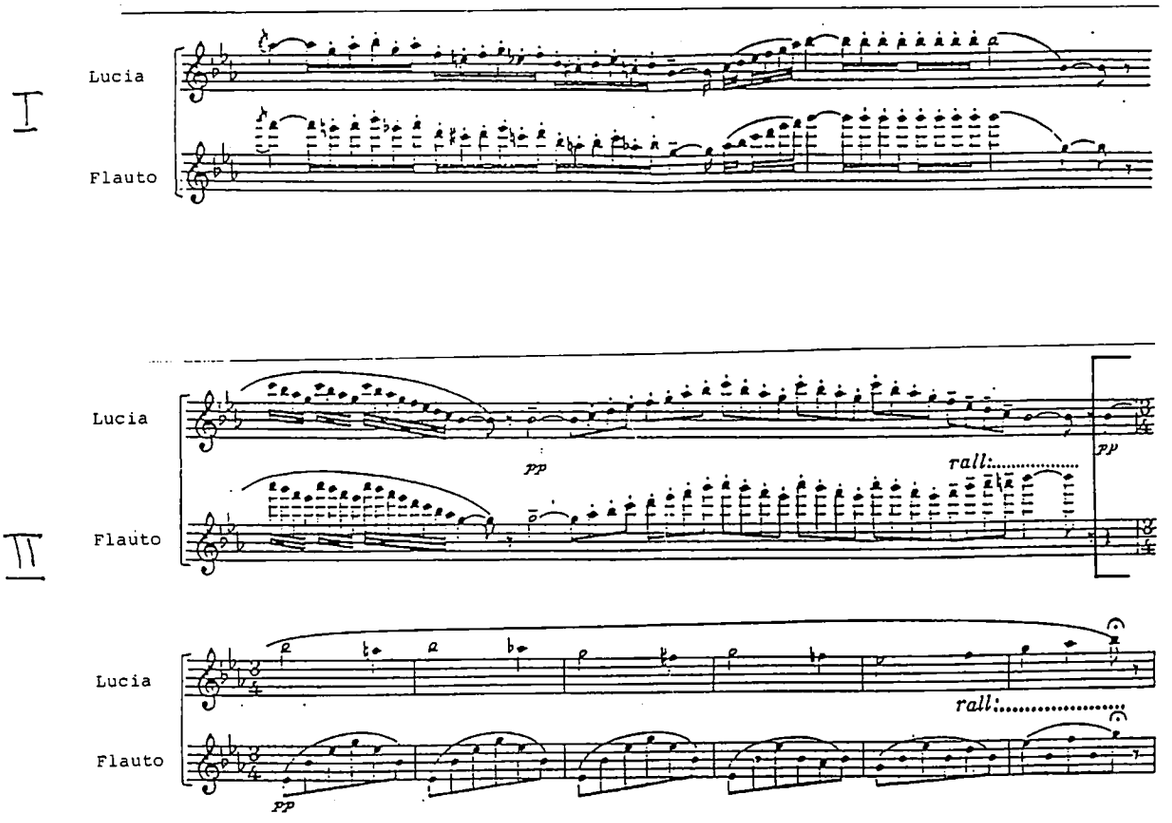

FIG. 1. Musical score of the two excerpts from the cadenza.

singer. It cannot be excluded that the recording engineers of the respective taping used various types of filter or equalizer circuits to differentially increase or decrease the energy in particular spectral regions. To ensure that there were no gross discrepancies in this respect between the different recordings, we analyzed the spectra of some pure flute notes from the instrumental parts of the cadenza. Judging from the frequencies of the partials as well as their respective energy relations, we did not discover major differences between the recordings. The exception is the historic Toti dal Monte record, which shows much noise throughout the spectrum. Therefore, the results for Toti dal Monte, reported later herein, should be treated as suggestive rather than conclusive. It should also be noted that the peak-picking routine used identified peaks on the basis of energy prominence only; thus, it is not possible to determine which types of peaks are being identified (formants, partials, etc.). Clearly, it would be desirable, in the future, to carry out much more precise analyses to determine the differential role of various types of peaks in the spectrum. Furthermore, it would be instructive to execute a dynamic analysis of the spectral changes to show to what extent the spectrum is affected by temporal phenomena, e.g., the local lengthening of a particular tone.

In any case, the potential differences in recording technique and resulting distortions of the spectral energy distribution do not directly affect the argumentation in the present article, which focuses on the role of acoustic singing cues—as mediated via a recording device—in determining the emotional *impressions* of listeners.

The excerpts were transferred to a Casio DAT recorder for the judgment study. A pairwise comparison design was chosen. For each of the two excerpts, all 10 possible pairs of the five singers were copied onto the tape, with an interstimulus interval of ~5 s. Each pair was separated by a pause of 25 s, during which time judges were to perform their ratings. To familiarize judges with the procedure, the judgment tape also contained three pairings of a brief excerpt from the role of Alisa (a minor character in the same opera, using three different singers performing the role: M. Di Stasio, Huguette Tourangeau, and Kathleen Kuhlmann).

A Casio DAT recorder with headphones was used for presenting the stimuli to judges. A standardized sound level, which had been determined in a pretest, was used throughout. Two different orders of the two excerpts chosen were used to balance out possible order effects.

The purpose of the judgment study was to obtain an indication of which emotion was attributed to a particular interpretation. The adaptation of the dynamics of emotional expression to the affective connotations of the text and the music, in line with the types of personal interpretation chosen, is one of the major prerequisites for a good opera singer. One of the methodological problems in choosing singers (on the basis of existing audio recordings) who differ greatly with respect to type of voice (e.g., dramatic versus lyric soprano), habitual timbre, and vocal technique, is that these voice-type-specific factors may influence listener judgment of emotionality. As a partial remedy, we also obtained judgments of preference, independent of emotional portrayal. It could be argued that although one might find great differences in preference for particular singers (depending on personal listening habits), there should be greater consensus on emotional expressiveness, if indeed a particular performance is marked by a specific emotional characterization (and if judgments on expressiveness are indeed independent of preference). In other words, one could assume that the singer's emotional involvement, or her professional ability to project a certain feeling without experiencing it during the performance, should be accessible to the listeners independent of vocal style or technique.

The instructions for the judgment task (in French) were provided on a cover sheet and a set of judgments scales were repeated for each stimulus on subsequent sheets. Based on the scene in which the cadenza occurs, we had chosen the following emotion scales for judgment:

1. Overall preference ("préférence générale")
2. Tender passion ("Désir ardent, bonheur, rêve d'amour")
3. Fear of death ("Angoisse de mort, effroi, saisie d'horreur")
4. Madness ("Folie, égarée, délirante")
5. Sadness ("Tristesse, mélancolie, désespérée")

We recruited 11 judges (5 male, 6 female), all of whom were interested in and knowledgeable about opera (to avoid effects of different degrees of familiarity with the art form of opera singing). Some of the judges were amateur or semiprofessional singers. Judges were to listen to each pair, indicate their overall preference for singer A or B in the pair, and

then judge for each of the four emotion scales which singer (A or B) expressed the respective emotion more clearly. Judges were encouraged to make a judgment even if they thought that the respective emotion was not really present in either of the portrayals. It was explained to them that judges are not always aware of their ability to make very fine distinctions. It was necessary to force the judges to make a decision each time, as is often the case in psychological judgment studies, because the pairwise comparison technique requires a complete set of judgments. In addition, judges had the possibility of writing comments on each pair of items.

RESULTS

Listener judgments

Our hypothesis was that although listeners might disagree about their overall preferences for a particular singer, they should tend to agree about the emotion(s) expressed in the respective interpretation. Furthermore, we were interested in seeing whether the different singers studied do in fact have differential affective impact on listeners.

We converted the binary judgments rendered for each pair into ranks by counting, for each judge, how often a particular singer had been chosen over all pairings and converting these numbers into ranks. If two or more singers reached the same number, tied ranks would be assigned. To test for agreement between judges, we recoded the ranks from 1 to 5 into three categories: Like (combining ranks 5, 4.5, and 4—i.e., the singer would always be first or second), Dislike (1, 1.5., and 2—i.e., the singer would always be last or second to last), and Indifferent for the intermediate range (3.5, 3, 2.5). We then computed a χ^2 for the cross-tabulation of the five singers with the three preference categories. Table 1 shows these results for the overall preference rating. The χ^2 of 51.68 is significant at the $p < 0.0000$ level, indicating that the differences in ranks between singers are not due to chance. In this case, this also indicates that judges agree more often than would be expected by chance in preferring one singer over another. Specifically, Gruberova is universally preferred, with 77% ranks in the Like category and very few ranks in the Indifferent or Dislike categories. Sutherland is next, with the ranks evenly distributed between the Like and Indifferent. Callas is quite special in that ranks are evenly distributed across all three categories, indicating large individual differences between judges in this case. Scotto has ranks mostly in the Indifferent and Dislike categories, Dal Monte virtually all in the Dislike category. It is not clear to what extent the result for Toti dal Monte is due to the fact that the historical recording used (which had been copied onto an audiocassette) did not match the auditory quality of the more recent recordings of the other singers that had been taken from CD (except Gruberova, in which a commercial MC was used). Some of the judges commented on the fact that the poor sound quality of the Dal Monte recording may have affected their judgment.

TABLE 1. *Overall preference*

		Like	Indifferent	Dislike	
Monte		2	1	19	22
	%	9.1	4.5	86.4	
Scotto		3	10	9	22
	%	13.6	45.5	40.9	
Gruberova		17	2	3	22
	%	77.3	9.1	13.6	
Callas		6	8	8	22
	%	27.3	36.4	36.4	
Sutherland		10	9	3	22
	%	45.5	40.9	13.9	

χ^2 51.6792, $p < 0.000$. Like = 1st or 2nd rank, Dislike = 4th or 5th rank, Indifferent = intermediate.

Rather than giving the detailed breakdown of the ranks for the emotion scales, we show the mean ranks for each scale and each singer in Fig. 2, thus summarizing the information. However, χ^2 analyses similar to the one in Table 1 were computed to establish the significance of the patterns (and evaluate the agreement between judges). Contrary to our hypothesis, agreement between judges on the emotions expressed seems to be somewhat lower than on overall preference. This is shown in the respective size of χ^2: Tender passion = 28.97, $p = 0.0003$; fear of death = 19.79, $p = 0.011$; madness = 10.58, not significant; and sadness = 39.72, $p < 0.0000$.

Thus, although the agreement on differences in the expression of emotion is still strongly above chance, it is less strong than in rating overall preferences for a particular artist. Looking at the pattern of data in Fig. 2, it seems that Gruberova (and to a lesser extent Sutherland) are seen as expressing more tender passion and sadness than the other singers. Sutherland is seen as particularly expressive of fear of death.

Given these patterns of data, one might inquire about the relationships between the preference

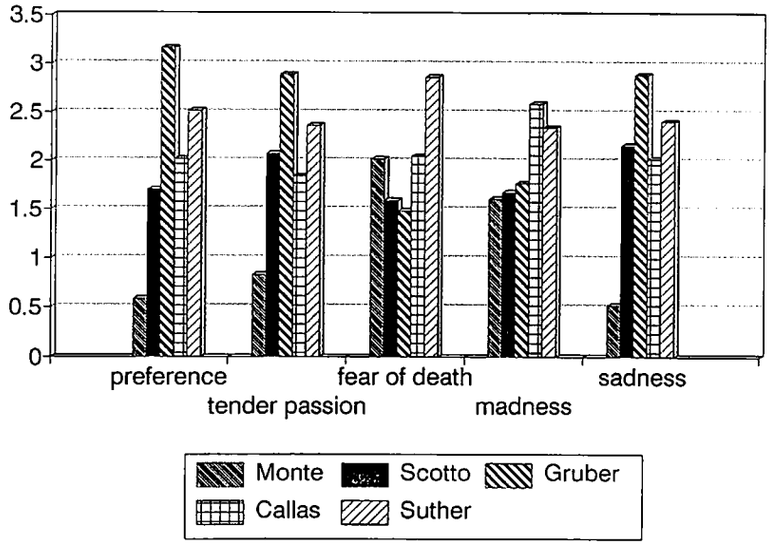

FIG. 2. Mean listener judgments—preference and emotional expression.

score and the emotion expression ratings. In fact, the Pearson correlation between the preference score and the emotion ratings (across all 10 stimuli from both excerpts[2]) reaches 0.97 for tender passion and 0.95 for sadness (−0.04 for fear of death and 0.34 for madness). One could argue, then, that listeners "automatically" assign high tender passion and sadness expression scores to their preferred singer. However, the relationships seem to be more complicated than are indicated by the correlations between the mean judgments. If one correlates overall preference and emotion expression for each individual listener, very complex patterns emerge: although the same pattern as found for the mean results is more or less present in six of the judges, for two others there are high correlations between preference and fear of death. No clear pattern is apparent for the remaining listeners. This suggests the possibility that listeners have idiosyncratic expectations about the nature of the emotion that is required of the role or in the specific scene and prefer the singer that seems to conform to the expectation. Although this finding may affect the generalizability of the findings, it does probably correspond to the reality of each listener bringing an individual frame of reference to bear in the appreciation of an artist's rendering of a particular piece.

The question as to how listeners proceeded in making their judgments is of interest in its own right. Did they first form an impression about affective expressiveness and decide on overall preference on that basis, or did they first decide on the interpretation they preferred and infer, on that basis, which emotion was likely to have been expressed (possibly being influenced by their personal notions about what emotions *should* be expressed in a "good" interpretation of that scene)? Unfortunately, we cannot expand on this interesting topic in the present context. We will take the mean ranks on all judgment variables as our dependent measure of audience appreciation for all further analyses.

Acoustic measurement

In this article we will not deal with dynamic variables of singing in the time domain such as tempo, pausing, vowel duration, phrasing, tone transitions, or similar phenomena that are highly variable in emotionally expressive singing. Given that we studied only two short excerpts without linguistic content, we thought that we did not have a sufficiently extensive sample for these variables. We therefore focused exclusively on acoustic measurements of voice quality (in the sense of spectral rather than temporal characteristics of the singing voice).

Individual variables

Although the acoustic characteristics were extracted independently for each of the two segments selected for study, we will present results averaged over the two segments to simplify the presentation. Because the acoustic profiles for the two excerpts

[2] This is not quite appropriate because the two observations per singer (the two excerpts) may not be completely independent because some subjects may have identified the singer.

per singer are very similar (because the voice characteristics of the singer seem to outweigh the differences in musical structure—at least with respect to the average spectrum), little information is lost. Although this is true for the restricted sample we looked at, in future studies one might want to look at the similarity of patterning in the vocal and temporal characteristics for the same singer across several different excerpts from one or more arias.

The means of all acoustic parameters measured for the five singers are listed in Table 2. Row 1 shows the average energy (intensity) of the voices. Row 2 the standard deviation of these energy values. Mean energy contains little diagnostic information because it partly depends on the recording level. However, it might influence listener judgment. The standard deviation of energy, on the other hand, is controlled by the singer. It partly reflects the variation in vocal effort during the two passages. Judging from the data, Gruberova and Sutherland vary intensity somewhat less than the other singers. It could be this very consistency that makes them the preferred singers.

Rows 3–10 show the average frequencies at which the automatic tracking program found energy peaks in the spectra and the standard deviation of the extracted energy peak values. It is not claimed that these peaks necessarily represent formant frequencies characteristic of specific vocal tract configurations. It is difficult to determine unambiguous formant frequencies for high-pitched voices, and our automatic analysis did not allow engaging in a microanalysis of likely formant frequencies. However, we will discuss various hypotheses as to the meaning of the peaks found by the peak-tracking algorithm. For each peak, first the mean frequency and then the standard deviation of the frequency values are shown.

The following attempts at interpreting these data should be taken with a grain of salt because it is obviously impossible to exclude the possibility that the sound engineers might have manipulated the relative prominence of different spectral regions (e.g., by using presence filters, etc., as mentioned in the Method section). This would particularly affect the mean frequencies of peaks, less the variabilities. In spite of these caveats, we decided to interpret the data with the aim of formulating hypotheses that can be tested in future studies in which it might be possible to exert better control on recording conditions. We were helped in this decision by the fact that the slight differences we found for the flute note spectra (see Method section) generally went in the opposite direction from the differences between singers.

Peak 1 most likely corresponds to the band for the fundamental frequency (F_0) of the voice (the notes prescribed in the score for the two segments of the cadenza range between b4 and b5, i.e., 488 and 976 Hz, with the majority in the upper region). Alternatively, for some of the tones the first peak might represent the second harmonic. Despite the

TABLE 2. *Mean acoustic variables for five singers*

	Dal Monte	Scotto	Gruberova	Callas	Sutherland
1 Total energy	77.64	67.70	65.52	69.15	71.28
2 Energy variability (SD)	16.84	13.84	12.17	13.45	10.12
3 Peak 1 energy	1162.82	1032.54	827.29	880.52	825.51
4 Peak 1 SD	263.53	301.33	120.36	137.98	127.44
5 Peak 2 energy	2351.14	2273.39	2522.54	2408.17	2161.70
6 Peak 2 SD	445.91	669.33	823.72	499.04	737.36
7 Peak 3 energy	3462.24	3398.40	3752.02	3469.82	3444.67
8 Peak 3 SD	817.51	687.90	753.98	345.99	474.53
9 Peak 4 energy	4477.07	4816.12	5825.34	4290.28	4844.82
10 Peak 4 SD	557.56	758.06	1880.17	498.15	773.30
11 Energy F_0 band	102.78	108.99	120.48	119.19	129.28
12 Energy harmonic 1 band	94.54	94.65	86.32	89.19	92.23
13 Energy harmonic 2 band	100.11	91.60	83.29	95.68	85.85
14 Energy formant 1 band	98.47	101.92	89.60	94.23	101.92
15 Energy singer's formant	98.66	89.94	78.77	96.54	88.52
16 Energy mid-frequency band	79.61	80.17	77.12	78.87	75.37
17 Energy high-frequency band	66.22	69.15	73.19	66.28	67.16
18 Comp Score 1 (F_0/Energy)	0.72	0.06	−1.47	0.68	0.02
19 Comp Score 2 (Sing.F/Hi Freq)	−0.55	−0.58	0.39	0.34	0.40

Rows 1–2, 11–17 in dB, rows 3–10 in Hz, rows 18–19 averaged z scores.

fact that the average should be about the same for all singers (as the notes are prescribed), there are noticeable differences. Scotto and particularly Dal Monte show rather higher F_0 band frequencies than the other three singers do. They also show higher variability of the peak in this band. Some of these differences might be due to the tuning of the respective orchestras. It is very difficult to determine what exactly peaks 2–4 represent. However, given the scarcity of data in this domain, we consider it useful to speculate about the possible significance of these empirically determined energy concentrations. As shown by Sundberg (15), the formant structure for vowels sung by sopranos changes quite dramatically depending on the fundamental. Because F_0 is very high in the present case, the first formant (F_1) is probably located at about the same frequencies for all singers. Sundberg (15, p. 126) has demonstrated that sopranos use the increased jaw opening to bring F_1 closer to the phonation frequency. Obviously, this is particularly effective for an /a/ vowel (or a schwa-like sound), as used in the cadenza, because the combination of F_0 and F_1 will produce strong energy.

The second formant for an /a/ vowel could be located at ~1,500 Hz for a fundamental of 700 Hz in a soprano (according to ref. 15, p. 126). At a mean F_0 level between 800 and 1,000 Hz as in the present case, one could imagine that F_1 is still higher. There is an energy peak in the average spectrum at ~1,800 Hz, which might correspond to F_2 or to the first harmonic, but this is not picked up by the peak tracking program, because it is quite a bit weaker than a peak at ~2,300 Hz, which is tracked as peak 2. Because (again according to Sundberg) F_3 tends to move down and F_2 tends to move up, our peak 2 might correspond to a blending of these two formants, which might have the function of a "singer's formant" at that range (see ref. 15, pp. 123–124, 142–143; and ref. 16, pp. 112–114; see later section herein).[3] The third peak at ~3,500 Hz could correspond to F_4, and peak 4 to F_5. Alternatively, if the sound produced during the cadenza is close to a schwa (unmodified vocal tract; with formants at 500, 1,500, 2,500, 3,500, 4,500 Hz), one could assume that all formant frequencies are shifted upward due to the shorter vocal tract for sopranos (see ref. 15, p. 131). This would fit relatively well with the overall peak structure that our tracking algorithm identified. The variability of the fourth peak might also reflect consonant production.

There is little difference between the singers for peak 2, although Gruberova has a slightly higher mean value and somewhat more variability of the peak frequency in this range. Although the third peak (and its variability) are relatively invariant across the singers, Gruberova again stands out with a higher fourth peak and much higher variability in this range. One might consider this to represent a quasi-singer's formant, accounting partly for the preference the listeners expressed for Gruberova. On the other hand, 4,500 Hz may seem very high, double the usual 2,200–2,300 area.

Although peak tracking allows one to determine at which frequency in the spectrum one finds energy concentrations, average spectrum analysis allows the determination of just how much relative energy is present at particular frequency ranges (which is, of course, partly dependent on recording conditions, including microphone placement, equalization, and other factors). We looked at seven theoretically specified bands in the average spectra for the two segments chosen: three F_0-related bands, two formant-related bands, and two mid-to-high-frequency bands. The delimitation of the bands was based on the investigation of the distribution of the respective parameters (F_0, harmonics) in the raw data for each singer. The first set is constituted by the F_0 band (~900–1,050 Hz, where the first peak—F_0—was located in the large majority of instances), a first harmonic band (~1,800–2,100 Hz), and a second harmonic band (~2,700– 3,150 Hz). One of formant bands is constituted by what we considered to be a probable range for the second formant of a schwa-like vowel (~1,550–1,750 Hz; see the foregoing discussion). The second formant band will be referred to as the singer's formant band (~ 2,300–2,600 Hz) even though it is debatable whether sopranos actually have a singer's formant. Sundberg (16, p. 113) cites a number of studies that find a much weaker singer's formant in sopranos than in other singers.

This is supported by a study reported by Watson (17). It seems that, in general, sopranos have far less energy in the high-frequency spectrum than do

[3] Although the phenomenon of a singer's formant between 2,500 and 3,500 Hz seems well established for male voices, there is debate about its existence for the female voice, particularly sopranos. However, Sundberg (15, pp. 124, 142–143) reports some evidence that female solo singers seem to have increased amplitude in the higher harmonic range and that the frequency of this region could be lower than in male singers. If this were indeed the case, our peak 2 at ~2,300 Hz might well represent such a female singer's formant.

male singers and altos. However, Sundberg (16) also shows that really famous sopranos, in comparison to a local singer, exhibited a rather sizable energy "hump" in the region between 2–3 kHz (ref. 14, pp. 113–114). Whether this energy concentration is functionally equivalent to a singer's formant is unclear. In any case, for our purposes of comparing the effect of different amounts of energy in specific bands of the spectrum, it does not matter whether this specific energy concentration functions as a singer's formant. The fact that, when singing with an orchestra, the best singers seem to produce as much energy in that region as they are capable of, indicates the utility of focusing on this spectral region for comparative purposes. For ease of reference, we will keep using the term singer's formant.

Finally, the upper part of the spectrum was divided into a mid-frequency (3,500–6,000 Hz) and a high-frequency (6,000–10,000) band. Because the latter is likely to consist mainly of low-energy noise, any differences found may be very unstable. However, we included this band in the analysis for exploratory purposes.

Rows 11–17 in Table 2 show that data for the standardized energy values for these frequency bands. The results for the F_0 and harmonics bands are interesting in that Gruberova, Callas, and Sutherland show much higher energy in the F_0 band (row 11) than do Dal Monte and Scotto. It may also be of interest that although the energy continuously drops from the fundamental (row 11) to the first (row 12) and then the second harmonic (row 13), as one would expect from a normal energy dropoff in the glottal spectrum, the second harmonic is actually stronger than the first for Dal Monte and Callas. With respect to the formant bands shown in rows 14 and 15, it is interesting to note that Gruberova shows rather less energy in the singer's formant range (row 15), particularly as compared with Dal Monte and Callas, in whom one finds a very strong energy component. Concerning mid and high frequencies (rows 16 and 17), the only difference to note is the relatively high energy level in the 6,000–10,000 Hz range for Gruberova (compared with the other singers).

Composite scales

Because there are sizable intercorrelations between the acoustic characteristics, we decided to form two composite scores [by taking the simple average of the standardized scores (z scores) for the cluster of interrelated variables]. The first score, which could be described as an F_0-band score, consists of peak 1 frequency (−), peak 1 variability (−), total energy variation (−), and energy in the F_0 band (+). Thus, a voice with a high score on this factor would have relatively low and stable F_0, with strong energy in the F_0 band and little total energy variation. The second score, a singer's formant/high-frequency factor, is related to the relative strength of the energy in the singer's formant band (~2,500 Hz) and the higher frequency ranges (between 3,500 and 10,000 Hz). The following variables were combined: energy in the singer's formant (+), high-frequency energy (−), peak 3 and 4 frequency (−), and peak 4 variability (−). Thus, a voice with a high score on this factor would have high energy in the singer's formant band but low energy in the high-frequency range as well as relatively low and invariant peaks in the 3,000–4,500-Hz range. Rows 18 and 19 in Table 2 list the respective values for the five singers.

Figure 3 shows how the five singers studied are located in the space formed by these two composite scores (which correlate with $r = -0.31$). Clearly, Gruberova is characterized by an extremely low position on one pole of the singer's formant/high frequency score with Scotto and Sutherland in a middle position and Dal Monte and Callas located toward the other pole of the dimension. On the F_0-band score, it is Dal Monte and Scotto who are on the opposite pole of the dimension with respect to Gruberova, Callas, and Sutherland. Given that only two very short samples were analyzed, these data cannot be used to infer stable differences in the voices of the five artists. Furthermore, with our present state of knowledge, it would be premature to attempt to speculate about the voice production and articulation techniques used on the basis of the acoustic parameters that were extracted. It should be noted that the parameters automatically extracted in our analysis may well miss some vocal features that contribute greatly to the emotionality of the interpretation. One such feature might well be vibrato, thought by Seashore (18) to be the only feature that matters in terms of emotionality. Clearly, future studies in this area should endeavor to include as many pertinent parameters as possible. On the other hand, it seems useful to determine to what extent classic acoustic parameters, relatively easily extracted from the singing voice, can help in understanding listener appreciation.

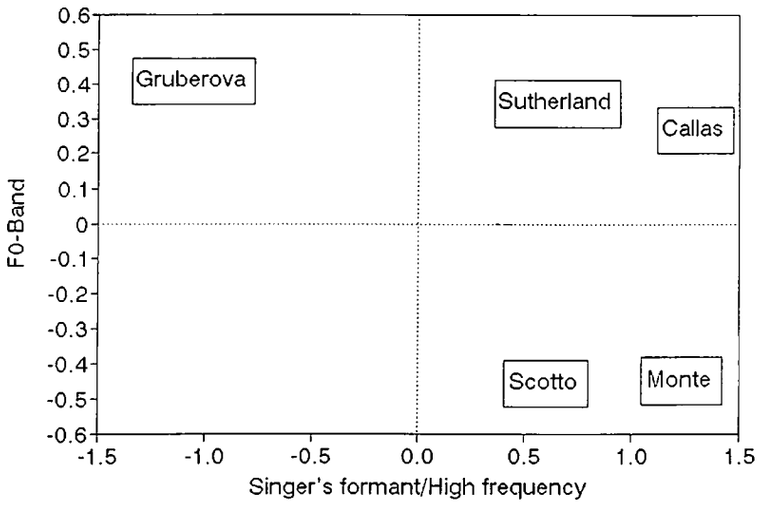

FIG. 3. Position of the five singers in the space formed by the two composite scores.

Correlations between listener judgments and acoustic characteristics

Although the acoustic data are hardly representative of the voices of the five "dive" studied, they should be a sufficient basis to determine some of the factors that influenced listener judgments. Clearly, because the two segments per singer were all the listeners heard, an objective acoustic description of these short segments does provide a representative account of the nature of the perceptual input. It is possible, of course, that listeners identified a particular singer and that, in consequence, stereotypical cognitive representations about that singer's performance influenced the ratings. However, we did not find that this happened frequently in our group of judges.

We correlated the mean ranks on the judgment scales and the acoustic characteristics over the 10 samples (two per singer).[4]

The correlations between the two composite scores and the mean ranks resulting from the listener judgments are shown in Table 3. The results are relatively straightforward: the second composite score—low energy in the singer's formant range and higher upper frequency peaks—seems to strongly influence overall preference judgment as well as tender passion and sadness ratings. Clearly, this result is largely due to Gruberova, who receives higher preference and tender passion/sadness ratings and who is placed in an extreme position on this score (obviously 2 points of 10 are likely to strongly influence a correlation coefficient). Although this is consistent with a sadness interpretation, in some sense, this finding is surprising with respect to the preference data. One would normally suspect that high energy in the 2,200–2,300 Hz singer's formant region produces high preference. It should be recalled that we found a strong energy concentration at double that frequency region for Gruberova, possibly a quasi-singer's formant with similar functional effects. If one excludes the values for Gruberova and computes a correlation over eight segments, the correlation for preference is only slightly lower (0.60), but the correlations for tender passion and sadness reach about the same levels. This indicates that the influence of the singer's formant/high-frequency score is more general and does not depend on one particular singer.

As shown by the correlations, the F_0-band score also influences subjective impression, although less strongly so. This is particularly true for overall preference—those voices with somewhat lower, stable F_0 and high energy in the F_0 range seem to be more appreciated.

To what extent can these acoustic variables ac-

TABLE 3. *Correlations between acoustic factors and judgments*

	F_0 band	Singer's formant
Preference	0.7	−0.78
Tender passion	0.53	−0.82
Fear of death	0.33	0.39
Madness	0.6	0.26
Sadness	0.53	−0.74

[4] This is again quite appropriate because the two observations per singer are not independent. However, given the small number of stimuli, we believe that it is better to base this exploratory analysis on a larger number of data points.

count for the variance in the judgment data? To answer this question, we regressed the two component scores onto the listener rankings. The results are shown in Table 4. Quite impressively, the two acoustic component scores explain 84% of the variance in the preference ratings. Explanatory power is somewhat lower for the emotion ratings, particularly fear of death, but nevertheless quite significant across all emotions. This demonstrates that the acoustic characteristics of the singers' voices that we identified and measured in this study contributed in a major way to listener appreciation. This may seem surprising because it is often held that energy flow and stability are the most significant factors that contribute to preference. However, the empirical data demonstrate the powerful contribution of the acoustic parameters analyzed in this study to the variance in the judgment data. One possibility is, of course, that at least some of the parameters we measured are closely correlated with dynamic energy flow.

CONCLUSIONS

Clearly, several aspects of the material used (rather brief excerpts, poor sound quality of the Dal Monte recording, recorded rather than live sound), which are likely to have affected both the listener judgments and the acoustic analyses, limit the generalizability of the results reported herein. However, with respect to the aims of the study as outlined in the introduction, we were able to show (a) that the different interpretations elicited significantly different listener ratings of emotional expressiveness (at least with respect to their judgments of expressed emotion), (b) that the voice samples of the five singers differ quite substantially with respect to objective acoustic variables (although the exact significance of some of these parameters still needs to be established), and (c) that we can quite successfully predict listener attributions on the basis of the objective acoustic characteristics.

TABLE 4. *Percent of the variance in listener judgment explained by two acoustic component scores*

Judgments	% explained (R^2)	F	p
Preference	0.84	18.50	0.0016
Tender passion	0.75	10.77	0.007
Fear of death	0.38	2.13	n.s.
Madness	0.59	4.94	0.046
Sadness	0.65	6.48	0.025

One might argue that we have not yet identified the "real" acoustic correlates of the emotional charging of the singing performances, where dynamic features might be highly important, and that the results given herein reflect perceptually irrelevant aspects of the recordings that covary with the perceptually crucial ones. Clearly, more detailed studies, allowing for a much higher degree of control of the pertinent variables, will be needed to investigate this possibility and to determine the role played by inflections and other dynamic acoustic gestures.

Admittedly, many of the differences found are due to the particular role of Edita Gruberova, both with respect to listener judgment (on both preferences and emotion attribution) and acoustic measurement; yet in a number of subanalyses we were able to show that the effects remain, albeit in weaker form, if we exclude Gruberova from the analysis. One reason for the special role this singer occupies in the group of "dive" compared here might be her voice type, coloratura of high soprano (a voice type that seems to be characterized by a particular timbre), as compared with the dramatic soprano voice types of most of the other singers in the group.

In spite of its limited generalizability, this study might serve to encourage further efforts to empirically tackle some of the questions raised by the emotional significance of music. Studying the expression of affective states in singing by using objective acoustic measurement techniques can help to link the study of emotional meaning in music to the literature concerned with the vocal expression of emotion in speech (8,11). Eventually, this might allow us to draw more conclusive inferences about the phylogenetic origins of speech and music (19).

More modestly, it would seem that the present results show that research paradigms currently used in the study of emotional expression in speech can be profitably used to look at emotion communication on the operatic stage. Clearly, though, some of the issues involved in opera singing may be much more complex than speech (including stage speech). One important issue, of course, is the simultaneous presence of music in many pieces, which makes acoustic analysis very difficult, if not impossible, and which does not allow the disentangling of the various factors. This, and the obvious technical problems in looking at recorded samples, make it seem imperative, in the long run, to work toward studies with a cappella singing in well-defined ex-

perimental paradigms. With the collaboration of artists, directors, and producers, it might well be possible to comprehensively study the processes of emotional encoding and decoding in a thoroughly empirical manner. Also, last but not least, important progress in the synthesis of the singing voice (20,21) would seem to soon allow study of the emotional effect of the singing voice by systematic variation of pertinent parameters via synthesis.

Acknowledgment: This study was conducted in collaboration with the Laboratoire d'Evaluation Psychologique at the University of Geneva. We thank Arvid Kappas and Ursula Scherer for helpful contributions to the acoustic and statistical analyses, and Ludovic Boyer and Christian Oberson for assistance in obtaining the judgments and preparing the data. We acknowledge valuable comments by Ivan Fonagy, Eric Clarke, and particularly Johan Sundberg, on an earlier version of the manuscript. We also acknowledge important comments by three anonymous reviewers whose suggestions have been partially incorporated into the manuscript.

REFERENCES

1. Kivy P. *Osmin's rage: philosophical reflections on opera, drama, and text*. Princeton: Princeton University Press, 1988.
2. Engel JJ. *Ideen zu einer Mimik*. Berlin: Mylius, 1785 (reprinted Darmstadt: Wissenschaftliche Buchgesellschaft, 1968).
3. Scherer KR, Banse R, Wallbott HG, Goldbeck T. Vocal cues in emotion encoding and decoding. *Motiv Emotion* 1991;15:123–48.
4. Wallbott HG, Scherer KR. Cues and channels in emotion recognition. *J Pers Soc Psychol* 1986;51:690–9.
5. Williams CE, Stevens KN. Emotions and speech: some acoustical correlates. *J Acoust Soc Am* 1972;52:1238–50.
6. Kotlyar GM, Morozov VP. Acoustical correlates of the emotional content of vocalized speech. *Sov Phys Acoust* 1976; 22:208–11.
7. Fonagy I. Emotions, voice and music. In: Sundberg J, ed. *Research aspects on singing*. Stockholm: Royal Swedish Academy of Music, 1981:51–79.
8. Scherer KR. Vocal affect expression: a review and a model for future research. *Psychol Bull* 1986;99:143–65.
9. Scherer KR. Vocal correlates of emotion. In: Manstead A, Wagner H, eds. *Handbook of psychophysiology: emotion and social behavior*. London: Wiley, 1989:165–97.
10. Frick RW. Communicating emotion: the role of prosodic features. *Psychol Bull* 1985;97:412–29.
11. Pittam J, Scherer KR. Vocal expression and communication of emotion. In: Lewis M, Haviland JM, eds. *Handbook of emotions*. New York: Guilford Press, 1993:185–98.
12. Scherer KR, Feldstein S, Bond RN, Rosenthal R. Vocal cues to deception: a comparative channel approach. *J Psycholinguist Res* 1985;14:409–25.
13. Cosmides L. Invariances in the acoustic expression of emotion during speech. *J Exp Psychol Hum Percept Perform* 1983;9:864–81.
14. Ladd DR, Silverman KEA, Tolkmitt F, Bergmann G, Scherer KR. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *J Acoust Soc Am* 1985;78:435–44.
15. Sundberg J. *The science of the singing voice*. DeKalb, Illinois: Northern Illinois University Press, 1987.
16. Sundberg J. What's so special about singers? *J Voice* 1990; 4:107–19.
17. Watson C. The singer's formant in the Scottish opera. *Bull Audiophonol* 1991;7:565–86.
18. Seashore C. *The psychology of music*. New York: Dover, 1967 (Orig. ed. New York: McGraw Hill, 1938).
19. Scherer KR. Emotion expression in speech and music. In: Sundberg J, Nord L, Carlson L, eds. *Music, language, speech, and brain*. London: Macmillan, 1991:146–56.
20. Sundberg J. Synthesis of singing. *Swed J Musicol* 1978;60: 107–12.
21. Risset JC. Speech and music combined: an overview. In: Sundberg J, Nord L, Carlson R, eds. *Music, language, speech, and brain*. London: Macmillan, 1991:368–79.