A singer’s expression of emotions in sung performance

Sundberg, J. and Iwarsson, J. and Hagegård, H.

journal: STL-QPSR
volume: 35
number: 2-3
year: 1994
pages: 081-092

http://www.speech.kth.se/qpsr
A singer's expression of emotions in sung performance

Johan Sundberg, Jenny Iwarsson & Håkan Hagegård

Abstract

A professional singer (co-author HH) sang a set of music excerpts from the Lieder and opera repertoires in two deliberately contrasting fashions, (1) as in a concert situation and (2) in an emotionally neutral way, as if the singer did not engage emotionally in the poem. The two renderings were evaluated by expert panels with respect to degree of emotional involvement and with regard to the type of emotional ambience characterizing the renderings. On the basis of these tests certain excerpts were selected for acoustic analysis. Surprisingly, no clear differences were found with regard to departures in nominal tone duration, thus suggesting that such departures are related to other aspects of expressivity than emotionality, such as the marking of musical structure. Agitated renderings seemed characterized by louder voice, faster tempo, and higher rate of loudness variation than the nonagitated renderings. Emotional expressivity seemed further associated with great F0 modulations during consonants and often also with a lowering of vowel formant frequencies.

Introduction

Emotional expression is an essential aspect of artistic singing; listening to a good performance can be as exiting as it is agonising to listen to a performance with a neutral or an inappropriate expression. Emotional expressivity also constitutes a highly relevant ingredient in speech communication. Yet, the expressive aspects of speech and particularly singing has attracted relatively little attention in research (see Scherer 1994 for a review). By and large, the results have shown that pitch characteristics have a very powerful effect on perceived emotionality in speech.

Studies of emotional speech do not explain how emotional expressivity is induced in sung performances, where the pitch parameter is restricted by the score. Some studies have been devoted to expressivity in singing. Kotlyar & Morozov (1976) asked eleven singers to sing a set of music excerpts imparting five different "emotional meanings" on the same music excerpts: joy, sorrow, anger, fear, and neutral. They measured syllable duration, micropauses between syllables, mean SPL, tone rise and decay time and found characteristic combinations in these measures for the various emotions. They concluded that temporal and dynamic characteristics are extremely important to vocal expression of emotions. Scherer & Oschinsky (1977) studied the relevance of amplitude modulation, tempo, harmonic content, tonality, rhythm as well as the level, contour and variation of pitch on emotion attributions to sentence-like sound sequences and musical melodies. The most important cues were found in the tempo and in the overtone content.

In this investigation we attempted to elucidate how emotions are expressed in sung performance by comparing deliberately neutral and expressive versions of the same
music examples. In this report some major trends will be described, leaving detailed analysis to future studies.

**Experiment**

An international baritone soloist, highly experienced in both opera and Lieder singing (co-author HH) agreed to perform a set of music examples in two different ways, with an expression which he perceived as appropriate and in a manner as void of emotional expression as possible. For obvious reasons, he experienced the latter task as somewhat exotic, lying beyond his musical expertise, though still possible.

The recordings were made on a DAT recorder in a large studio hall at the Hagegården Music Center, Gunnarsbyn, Sweden. A small microphone was fastened to the singer's glasses so as to ensure a constant distance. No piano accompaniment was used.

The excerpts are listed in Table 1. They were selected so as to potentially represent a variety of different emotional ambiances, such as fear, joy, sadness, love, hate, security, tenderness. The singer was not informed about these emotional interpretations of the excerpts.

<table>
<thead>
<tr>
<th>Composer</th>
<th>Song title</th>
<th>Excerpt Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Folktune</td>
<td>&quot;Byssan lull&quot;, verse 1</td>
<td>Byssan</td>
</tr>
<tr>
<td>Folktune</td>
<td>&quot;Vi gå över daggstänkta berg&quot;, bars 1-8</td>
<td>Daggstänkta*</td>
</tr>
<tr>
<td>G. Mahler</td>
<td>&quot;Lieder eines fahrenden Gesellen,&quot; song # 3, bars 5-11</td>
<td>Glühend Messer*</td>
</tr>
<tr>
<td>F. Mendelsohn</td>
<td>&quot;Paulus&quot;, Aria # 18, bars 5-13</td>
<td>Mendelsohn</td>
</tr>
<tr>
<td>F. Schubert</td>
<td>&quot;Erlkönig&quot;, bars 41-50</td>
<td>Siehst Vater Du*</td>
</tr>
<tr>
<td>F. Schubert</td>
<td>&quot;Erlkönig&quot;, bars 72-79</td>
<td>Mein Vater*</td>
</tr>
<tr>
<td>F. Schubert</td>
<td>&quot;Du bist die Ruh&quot;, bars 8-15</td>
<td>Du bist die Ruh</td>
</tr>
<tr>
<td>F. Schubert</td>
<td>&quot;Wanderers Nachtlied&quot;, bars 3-14</td>
<td>Wanderers Nachtlied</td>
</tr>
<tr>
<td>F. Schubert</td>
<td>&quot;Das Wandern&quot;, verse 1, bars 12-20</td>
<td>Das Wandern</td>
</tr>
<tr>
<td>F. Schubert</td>
<td>&quot;Nähe des Geliebten&quot;, bars 3-8</td>
<td>Nähe des Geliebten</td>
</tr>
<tr>
<td>R. Schumann</td>
<td>&quot;Liederkreis&quot; VI &quot;Schöne Fremde&quot;, bars 16-24</td>
<td>Es funkeln *</td>
</tr>
<tr>
<td>R. Schumann</td>
<td>VIII &quot;In der Fremde&quot;, bars 25-37</td>
<td>Als müsste*</td>
</tr>
<tr>
<td>R. Schumann</td>
<td>XII &quot;Frühlingsnacht&quot;, bars 18-26</td>
<td>Und der Mond*</td>
</tr>
<tr>
<td>R. Strauss</td>
<td>&quot;Zueignung&quot;, bars 21-29</td>
<td>Es schweben</td>
</tr>
<tr>
<td>G. Verdi</td>
<td>&quot;Falstaff&quot;, Ford's Monologue, bars 24-31</td>
<td>Wie Du auch*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zueignung*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ford's Monologue*</td>
</tr>
</tbody>
</table>

**Perceptual evaluations**

Two listening tests were carried out to evaluate the emotional quality of the performances recorded. In test 1 the excerpts were arranged in random order on a tape.
Using loudspeakers, this tape was individually presented to six highly experienced experts on sung performance. They were asked to show by a mark on a 10 cm long line how much expressivity they heard in the different performances. Before the proper test, a set of six examples was presented so as to familiarise the subjects with the task.

There were five replicated stimuli on the tape; the mean rating difference for these pairs ranged from 0.7 to 2.3 cm, intersubject mean 1.4 cm, thus suggesting that the subjects were rather consistent. The mean standard deviation for the subjects' ratings of identical stimuli also amounted to 1.4 cm, thus indicating a fair degree of agreement. The ratings were averaged and the five excerpts which showed the smallest mean rating difference between neutral and expressive and/or a low rating of the expressive version were eliminated from further analysis. This procedure yielded paired versions of 12 excerpts.

This test helped us select those pairs of versions which differed clearly in expressivity, but did not indicate the nature of this expressivity difference. To elucidate this, a second test was carried out. A test tape was prepared in which the renderings were arranged in pairs, each presented twice in succession. In each pair, the first version was the neutral one and the second the expressive one. A pretest pair was presented first on the tape to familiarise the subjects with the procedure. In this test the task of the subjects was to decide if the versions differed in any of seven standard emotions: secure, loving, sad, happy, scared, angry, and hateful. By marks in boxes, one for each emotion, the subjects indicated if they thought that the singer sounded more sad, happy, scared etc. in the expressive version. The subjects were five highly experienced singing teachers.

The results of the perceptual and acoustical analyses are summarised in Table 2. In the rightmost column are listed the emotions which received at least three votes for the expressive version in listening experiment 2, the numbers within parentheses indicating the number of votes. As shown by the table, one or more of the seven emotion alternatives seemed applicable for describing aspects of the difference between the versions in most cases. The emotions loving and secure turned out to be frequently represented. In no case was there a predominance of one single emotion. This suggests that music expression is hard to describe by words representing one single emotion.

**Acoustic analysis**

The tempo, defined as the inverted mean duration of the shortest note value present in the excerpt, is listed in Table 2. In most cases the tempo was slower in the expressive than in the neutral versions. The differences were small for the excerpts where the expressive version was characterised as hateful, angry, happy, thus suggesting that the singer did not signal these emotions by great tempo effects. On the other hand, the emotions secure and loving were all associated with a clearly slower tempo in the expressive version. A slightly faster tempo was observed in one case, only, the expressive version of Dagsstänkta, characterised as happy. This supports the observation by Gabrielsson (1993) that happy is associated with a fast tempo.
Table 2. Overview of results of acoustic analysis of the differences between the expressive and neutral versions of the excerpts. For each parameter, the values observed for the neutral versions are given together with the difference in percent from this value in the expressive version. The tempo is given as the number of the excerpt's shortest note value (SNV) per second, the loudness in terms of the sound level (SL) averaged over the entire excerpt, and the short-term variability of loudness (Mean dSL/dt) in terms of the time derivative of sound level averaged over the entire excerpt. The leftmost column shows the predominant emotions of the expressive versions according to Listening test 2. The excerpts are ordered according to the loudness variability value. Excerpts marked with a star (*) possess an agitated or lively musical character.

<table>
<thead>
<tr>
<th>Excerpt</th>
<th>Tempo</th>
<th>Mean SL</th>
<th>Mean dSL/dt</th>
<th>Predominant emotion (&lt;3 of 5 votes)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neutr</td>
<td>Expr</td>
<td>Neutr</td>
<td>Expr</td>
</tr>
<tr>
<td></td>
<td>SNV/s</td>
<td>Diff (%)</td>
<td>(dB)</td>
<td>Diff (%)</td>
</tr>
<tr>
<td>Ford's Monologue *</td>
<td>6.1</td>
<td>-4.9</td>
<td>75.9</td>
<td>-0.7</td>
</tr>
<tr>
<td>Mein Vater *</td>
<td>5.7</td>
<td>0</td>
<td>70.7</td>
<td>2.5</td>
</tr>
<tr>
<td>Ich hab ein *</td>
<td>5.8</td>
<td>-4.3</td>
<td>76.9</td>
<td>2.6</td>
</tr>
<tr>
<td>Wie du auch *</td>
<td>8.7</td>
<td>-6.9</td>
<td>69.6</td>
<td>3.2</td>
</tr>
<tr>
<td>Und der Mond *</td>
<td>4.0</td>
<td>0</td>
<td>59.3</td>
<td>2.7</td>
</tr>
<tr>
<td>Zueignung *</td>
<td>5.3</td>
<td>-16.9</td>
<td>71.7</td>
<td>-0.2</td>
</tr>
<tr>
<td>Daggstänkta *</td>
<td>7.9</td>
<td>3.8</td>
<td>55.8</td>
<td>1.7</td>
</tr>
<tr>
<td>Nähe des geliebten</td>
<td>6.8</td>
<td>-26.5</td>
<td>75.8</td>
<td>-1.2</td>
</tr>
<tr>
<td>Mendelsohn</td>
<td>4.5</td>
<td>-20.0</td>
<td>77.6</td>
<td>0.8</td>
</tr>
<tr>
<td>Es schweben</td>
<td>8.0</td>
<td>-14.0</td>
<td>78.5</td>
<td>-1.2</td>
</tr>
<tr>
<td>Wanderers Nachlied</td>
<td>2.6</td>
<td>-23.1</td>
<td>73.2</td>
<td>-1.3</td>
</tr>
<tr>
<td>Du bist die Ruh</td>
<td>2.5</td>
<td>-20.0</td>
<td>76.8</td>
<td>-4.5</td>
</tr>
</tbody>
</table>

Vowel-to-vowel durations were measured from spectrograms and compared with the score's nominal durations, defined from the data on the mean tempo. The departures from the nominal durations tended to form patterns that were mostly similar for the neutral and expressive durations, a finding nicely replicating results from an experiment recently reported by Drake & Palmer (1993). These patterns seemed to reflect two principles, lengthening of short notes, and marking of musical phrase structure. This marking adhered to a principle previously described by Todd (1985); a phrase is typically marked by starting somewhat slowly, then accelerating, and again decelerating towards the end of the phrase, so that a parabolic tempo curve is produced, see Figure 1a. Applying these two principles to the nominal durations of two of the excerpts yielded an astonishingly good match of the singer's performances in both the neutral and the expressive rendering, as can be seen in Figure 1b and c. Thus, independent of emotional expressivity, the singer's marking of phrases can be rather accurately described using the same coefficients. Greater lengthenings of short notes were sometimes observed in the expressive versions. Somewhat surprisingly, no examples were found of lengthening of notes presenting particularly important words in the poem.
Figure 1. Upper panel: Parabola pattern used for increasing the durations (DR) of notes in a subphrase and a phrase (left and right half of the graph), see Todd (1985).

Middle and lower panels: Deviations from nominal durations (DR) observed for the various syllables in excerpts "Es schweben" (middle) and "Mendelsohn" (lower). Open circles and filled diamonds refer to the neutral and expressive versions, and the open diamonds and dotted lines refer to model generated data, using the parabola pattern shown in the upper panel and the principle of lengthening of short notes.
Overall sound level was determined as the mean sound level over the entire excerpt as measured by the SWELL program (Ternström, 1991). The means are given in Table 2. Note that this mean increases not only when vocal loudness is high, but also when a song is performed legato, with few and short intertone pauses. The latter effect is reflected in the low value obtained for Daggstänkt, which the singer sang in a staccato-like manner. Expressive versions characterised as secure and loving showed lower means than in neutral, as they were sung softer. The opposite effect was noted in most versions characterised as hateful or happy.

Comparison of the long term variations of sound level between the neutral and expressive renderings revealed no clear differences, thus suggesting that crescendos and diminuendos were not used for the purpose of imparting emotional expression. In excerpt "Und der Mond", however, the singer made a crescendo at the end of the piece only in the emotive version. This probably reflected the content of the poem; in this part of the song the text presents the essence of the poem. Thus, in this case the singer seemed to use long term variation of sound level for the purpose of enhancing aspects of the text. The short-term variability of sound level was evaluated by computing the time derivative of the sound-level curve, smoothed with a 20 Hz low-pass filter. The means of this derivative are listed in Table 2, where the excerpts were ordered according to this mean. The difference in this parameter was considerable, ranging from +93% to -50%. Increases were found in the expressive versions characterised as hateful or happy while decreases were found for most secure and loving versions. Indeed, all agitated excerpts were found at the upper part of the table while all excerpts portraying a nonagitated emotion are found in the lower part.

The source of a great variability was mostly an emphatic pronunciation of consonants causing a great amplitude modulation in the expressive versions. Figure 2 offers a typical example, where in the expressive version examples of a great amplitude modulation can be observed, e.g., during the consonant [r] in "strahlstl". The above observations show that the short-term variability in sound level is an important means of emotional expression in singing.

F0-patterns were analysed by means of the double-peak-picking strategy available in the SWELL program. In agitated excerpts, the extent of the vibrato undulations was mostly considerably greater in the expressive than in the neutral versions, while in the nonagitated excerpts the two versions were more similar. Typical examples are shown in Figure 3. This suggests that the vibrato extent is used for expressive purposes in both singing and instrument playing (Gabrielsson, 1993).

Local departures from the nominal F0 values were observed in many expressive versions. These departures often smoothed pitch jumps by means of short-term portamenti, i.e., glissando-like pitch changes, but sometimes F0 simply made a deep dip between tones so that the new tone onset started from a very low value. Such portamenti frequently occurred on voiced consonants, apparently for the reason of emphasis.
Figure 2. Comparison of audio signal, sound level and F0 in the neutral (upper panel) and expressive (lower panel) versions of the first tones in excerpt "Wie du auch strahlst". Note the amplitude modulation in the first three syllables and during the [r] sound in the expressive version.
and a monofrail (Wanderers Nachtlieder, lower) except for the difference in vowel.

Figure 3: F0 patterns for neutral and expressive versions of an agglutinated (Wanderers Nachtlieder)
According to listening experiment 2 the expressive version collected all subjects' votes for one or more specific emotions in five excerpts (see Table 2). Using spectrographic analysis (Sonagraph DSP 5500) the four lowest formant frequencies were determined for most vowels in these excerpts. Figure 4 compares the data for the neutral and expressive versions. Occasionally differences were observed for F1 and F2 and in all such cases the value measured in the emotive rendering was lower than that measured in the neutral rendering. For F1 and F2 the differences occurred in the ranges of 0.5-0.8 kHz and 1-2 kHz, respectively. Hence, the lower values in the emotive versions reflected a darker vowel quality. F3 and in particular F4 frequently showed clearly lower values in the emotive renderings. This would mirror a "coloring" of the voice timbre which would correspond to a more enhanced singer's formant. The result suggests that voice timbre can be used for the purpose of emotional expression in singing.

\[ F1, F2, F3 \text{ & } F4 \]

Figure 4. Comparisons of F1, F2, F3 and F4 of vowels occurring in the neutral and expressive versions of five of the excerpts. Symbols refer to formant numbers.

Discussion
In this investigation we attempted to use as musically realistic situations as possible, thus taking advantage of the singer's special competence. We avoided asking him to perform the same excerpt with a number of different emotions, because the musical competence of a musician would include the ability to portray by acoustical means the
particular emotional ambiance embedded by the composer in the song while dressing a
song in an inappropriate ambiance is not necessarily part of a musician's expertise.

Our results revealed two major performance trends. One was the marking of the
musical structure by means of patterned departures from nominal duration. These
departures were rather similar in the neutral and expressive versions. The second trend
was to use tempo, fundamental frequency modulation, loudness and rate of loudness
variation for expressive purposes. These factors seemed to vary in accordance with the
overall musical character of the excerpt.

As mentioned before, the excerpts can be grossly divided into two groups, agitated
and nonagitated. The tempo tended to be considerably slower in the expressive
renderings of the nonagitated excerpts. With regard to loudness, most expressive
versions of the agitated excerpts were sung louder while those of the nonagitated
versions were sung softer.

With regard to rate of loudness variation, expressive versions of all agitated excerpts
showed increased rates while those of the nonagitated excerpts showed decreased rates.
These observations suggest that the polarity agitated/nonagitated is a factor of major
relevance to expression in singing. However, signs of specific emotions may very
well hide in small details, such as formant transitions at CV and VC boundaries.
Synthesis would offer a powerful tool in revealing the emotional relevance of such
details in future research.

By and large, our observations agree with findings made by Kotlyar & Morozov
(1976). For instance, they found that sad was characterised by long durations of
syllables, i.e., a slow tempo, that the "average voice pressure level" was high in anger
and also in happy, and that syllable onsets and decays were fast in anger. Furthermore,
our results reveal striking parallels between expression in speech and singing. Scherer
(1994), summarising previous research of expression of emotions in speech, notes that
anger and joy are associated with a high F0 variability and also that sadness is
associated with a low vocal intensity and a low rate of articulation. Hypothesising that
F0 variability in speech is translated into vibrato extent in singing, these observations
seem to compare well with our findings.

The departures from nominal duration showed that short notes were often
lengthened in both the neutral and the expressive versions. Previous studies of musical
performance of instrumental music has revealed the opposite trend, i.e., that short notes
are shortened (Sundberg, 1993), but a detailed analysis did not show one single
example of this principle in the singer's expressive versions, not even in the agitated
excerpts. Lengthening of short notes may be idiomatic to singing, although singers and
instrumentalists mostly perform music together and can hence be expected to adhere to
similar performance principles.

The relevance of the text to the artist's sung performance is interesting. Palmer and
Kelly (1992) analysed syllable durations in sung performances and observed greater
lengthenings of nouns than adjectives under certain conditions. We did not find any
cases of lengthening of semantically important syllables in our material. It might be
rewarding to study such aspects further by comparing a singer's sung performance with
an actors' reciting of the same text. It would also be interesting to compare expressive
performances of songs sung with its text and as a vocalise, i.e., as a song with sustained vowels replacing the text.

One single singer was used in the present investigation, so the results do not allow general conclusions. On the other hand, the fact that this singer is successful in an international career obviously makes his methods of expression a worthwhile object of scientific study. The singer's success also suggests that his methods of conveying expression in singing are widely accepted. Moreover, our study revealed striking parallels between expression in speech and singing. Still, we believe that some characteristics of his art may be personal. For example, it was interesting that in listening test 2, there was a predominance of loving and secure, while scared and sad were rarely chosen alternatives.

This investigation was a preliminary attempt to study emotional expression in singing, and our findings have resulted in suggestions for a number of future studies which could complement the description of emotional expression in singing. Thereby, we think that our method of comparing neutral with expressive versions of the same excerpt is useful. The generality of the findings should be tested in synthesis experiments and by analysing other singers' methods for expressing emotions.

Conclusions

An agitated versus nonagitated overall character of the song seems to be highly relevant to the expression of emotions in singing. Fast tempo, high vocal loudness and, in particular, high rate of loudness variation were used in agitated excerpts while the opposite was observed in nonagitated excerpts. By and large, these characteristics agree with those previously found in speech. Expression seemed further promoted by great fundamental frequency modulations during consonant segments, and sometimes also by a lowering of the vowel formant frequencies. Departures from nominal tone durations, on the other hand, seemed rather independent of emotional expression, apparently depending more on the musical structure. A more detailed analysis, preferably supported by synthesis experiments, is likely to further elucidate the expression of emotions in singing.

Acknowledgement

The authors gratefully acknowledge the participation of Lars Frydén and Anders Friberg in the analysis of the singer's departures from nominal duration. The research was supported by a grant from the Swedish Research Council for Engineering Sciences.

References


