Published paper

Timmers, R. (2007) *Vocal expression in recorded performances of Schubert songs*, Musicae Scientiae, 11 (2), pp. 237-268

Vocal expression in recorded performances of Schubert songs

Musicae Scientiae XI (2), 237-268.

Renee Timmers[a]

Centre for the Analysis and History of Recorded Music, Department of Music

King's College London

Strand, London WC2R 2LS, UK

r.timmers@sheffield.ac.uk

a) Current affiliation: Department of Music, University of Sheffield

Abstract

This exploratory study focuses on the relationship between vocal expression, musical structure, and emotion in recorded performances by famous singers of three Schubert songs. Measurement of variations in tempo, dynamics, and pitch showed highly systematic relationships with the music's structural and emotional characteristics, particularly as regards emotional activity and valence. Relationships with emotional activity were consistent across both singers and musical pieces, while relationships with emotional valence were piece-specific. Clear changes in performing style over the twentieth century were observed, including diminishing rubato, an increase followed by a decrease of the use of pitch glides, and a widening and slowing of vibrato. These systematic changes over time concern only the style of performance, not the strategies deployed to express the structural and emotional aspects of the music.

1. Introduction

Recordings form a rich source of information about musical performances from the past, and are especially indispensable as a source for histories of performing style (*e.g.* Philip, 1992); Day, 2000); Fabian, 2003). These investigations have highlighted changes in performance characteristics such as tempo, rubato, or the use of vibrato and glissandi in singing and string performances. For example, a change in attitude towards rubato was observed in the first half of the twentieth century (Philip, 1992; Hudson, 1994; Brown, 1999): tempo fluctuations in recorded performances show a trend related to this change in attitude from frequent tempo changes to smaller and gradual tempo modifications. Another example is that vibrato came to be seen as part of the sound colour rather than as an ornament that embellishes certain notes (Philip, 1992; Brown, 1999). This resulted in more continuous and more prominent vibrato, applied even to notes of short duration, increasing their power and penetration.

Historical recordings are less often used to investigate local variations related to performers' interpretations of the structure and emotion of specific pieces. Relationships between performance variation or "expression" and structural or emotional characteristics of the music are more usually investigated in empirical studies of contemporary performers; for example, Sloboda (1983) measured variations in piano performances related to different interpretations of metrical structure, while Palmer (1989) examined systematic changes in piano performances related to different interpretations of phrase structure. And Gabrielsson and Juslin (1996) examined expressions related to different emotional interpretations in instrumental performances. These studies have highlighted "strategies" of communication, such as systematic variations in tempo and dynamics to communicate phrase structure or a systematic relationship between choice of tempo and communicated emotion.

The aim of the present study is to combine the investigation of performance strategies with a historical perspective and to apply this to the analysis of vocal performances on record. The study specifically investigates the relationships between vocal expression, musical structure and emotion, and how these relationships have changed during the twentieth century in performances of three Schubert songs. It takes an analysis by measurement approach: vocal expression is measured from modern and historical recordings and related to coded descriptions of the music.

When analysing natural performances from records, one should consider multiple possible sources for performance variations. The use of a model is often helpful to investigate complex data in a systematic manner. For example, Juslin and colleagues (2002) suggested a model for expression called GERM. The title refers to the components of the model, which include generative expression of structure, expression of emotion, random fluctuations, and analogies to biological motion (such as their model of the final ritard). The expression model can be used to create performances, but also to relate measured performance variations to different sources of variation. Windsor and colleagues (2006) have similarly argued that performance variations may be decomposed into constituents by using a detailed description of the musical structure. They demonstrated the success of such an approach for timing variations in piano performances of a simple Beethoven piece at different tempi. A linear relation between aspects of the musical structure and performance variations was assumed.

These methods of performance analysis are based on the assumption that an important role of performance is to communicate the structure and emotion of music to listeners (Palmer, 1997; Gabrielsson, 1999). Moreover, performers seem to generate expressive variations from an internal representation of the structure of music (Sloboda, 1983; Clarke, 1988).

The model used in the present study is based on the assumption that expressive variations in performance variables are related to the structure and emotion of the music. In other words, it takes two of the components of the GERM-model into account and neglects the other two. A regression model is built from an analysis of structural and emotional characteristics of the songs, as explained in detail in the method section. This procedure is suited for highlighting relationships between performance variations and pre-defined musical characteristics. It may highlight strategies for communicating these musical characteristics by different means: for example, dynamics or tempo may be used to communicate phrase endings, while variations in vibrato may (or may not) be used to communicate musical structure. Multiple regression analysis may also highlight differences in emphasis of musical characteristics; singers may show systematic relationships between performance variations and different aspects – structural or emotional – of the music.

To use recordings and measure performance parameters as duration, amplitude and pitch is tricky. Recordings are not faithful representations of live performances, partly because performers may have adapted their manner of performance to the recording conditions, partly because recording conditions (including equipment) modify the sound being recorded, and partly because playback equipment interprets information on the disc. The original playback speed of a recording (especially early recordings) is uncertain. Differences in playback speed influence measurements of duration and vibrato rate as well as measurements of pitch. This suggests that relative variations in tempo and pitch within a recorded performance are more reliable than the average speed or pitch. However even these variations within a recorded performance may have been modified, from the LP era onwards, by splices and cuts of several takes.

Therefore the interpretation of the results should proceed with caution, and with knowledge of the history of recording (ideally of each recording used). The results reflect the characteristics of recorded performances: not only the singer and accompanist are responsible for these characteristics, but a sound engineer and several items of equipment as well. On the other hand, the recording certainly has validity as a sound object that is "real" in the sense that it is listened to and used as a commodity. Below, the analyses of these sound objects are reported without pondering the details of how recordings at different time periods came into existence. Interested readers are referred to a growing literature on contexts and techniques of recordings (see the CHARM website, www.charm.rhul.ac.uk; Day, 2000; Leech-Wilkinson, 2007).

Before the details of the method are explained, relevant literature is first summarized on relationships between performance expression and musical structure and emotion. This literature provides the main background for the presented investigation of vocal expression.

2. Relationships between expression, structure and emotion

As mentioned above, the assumption is that an important aspect of music performance is to communicate the structure and emotion of music to listeners. Simple relationships between expression and musical structure are assumed, such as slowing down at the end of a phrase, increase in contrast between short and long notes, speeding up when the melody rises in pitch, or intensification with harmonic charge, which is a measure of tonal distance (*e.g.,* Sundberg *et al.,* 1989; Todd, 1992). In the current study, previous studies are followed and descriptions of phrase structure, rhythmic density, melodic contour, and melodic highpoint are included. No descriptions of harmony are directly included: instead harmony is used in the characterization of the emotional characteristics of different passages.

Previously observed relationships between performance variations and emotions in music include those related to expressions of basic emotions, such as happiness, sadness, fear and anger, and to expressions of emotional valence, activity and tension (discussed below). Expressions of basic emotions in singing and instrumental performances have been quite extensively investigated (*e.g.,* Sundberg, 1987; Scherer, 1986, 1995; Ohgushi & Hattori, 1996, Juslin & Laukka, 2003). A close link between expression of emotions in music and speech has been suggested. This connection is evident from similarities in acoustic characteristics associated with expressions of specific emotions. For example, anger is expressed by sounds larger and stronger than normal: fast, loud, fast attack, high energy, and high variability in sound level. Expressions of fear tend to be fast, low in loudness, but with high variability in sound level. Happiness is expressed using high tempo, medium loudness, stable sound, fast attack, and upwards pitch contours (Juslin & Laukka, 2003).

A difficulty related to the expression of emotions in music is that it is almost never entirely clear which emotion is expressed. Listeners may categorize the emotion perceived in music consistently when asked to select from a limited number of possibilities, such as four to six basic emotions, but when confronted with a larger number of choices agreement tends to decrease (Gabrielsson & Juslin, 2003). Moreover, not all authors agree on the prevalence of basic emotions, either in the general psychological literature (Ortony & Turner, 1990; Scherer *et al.*, 2003), or in that dealing with music. Emotional valence and activity (or arousal) are sometimes used as alternatives for characterizing the emotion perceived in music (Schubert, 2001, 2004; Scherer, 2004), as they capture the relationships between emotions well (Russell, 1980): subjective distances between emotions tend to correspond to the distance between emotion descriptors along these two dimensions. Emotions may be positive or negative, implying approach or withdrawal, such as happiness compared to fear: this affective direction is referred to as the valence of emotions. Additionally, emotions may be high or low in arousal (referred to as emotional activity). For example, anger has a high level of arousal, while depression and boredom have low levels of arousal.

However, not all theorists agree that music expresses different emotions. Emotions represented in music tend to continually increase or decrease in intensity, for some theorists, this pattern of ebb and flow or tension and relaxation is specifically characteristic of emotions in music (*e.g.*, Langer, 1942; Meyer, 1956). Related empirical studies have asked listeners to indicate changes in tension or emotional intensity by moving a slider while listening to music. These indications of intensification and abatement tend to correlate with the dynamics and tempo of the performance (Schubert, 2001; Timmers *et al.*, 2006), and with moments of introduction of new material and closure (Krumhansl, 1996), as well as with moments marked by the performer (Sloboda & Lehmann, 2001).

An assumption of the present study is that most passages in Schubert songs have a prevailing emotional characteristic determined by a combination of the text and the music. Although musical passages may be possibly ambiguous or have several layers of meaning, a prevailing emotional characteristic in terms of valence and activity can often be determined. An additional assumption is that tension tends to fluctuate within music passages. Changes in emotional qualities are examined at two levels: 1) at the level of musical passages; and 2) at the bar level within musical passages. The former are characterized in terms of emotional valence and activity, and the latter in terms of local highpoints in tension. This procedure is in line with previous empirical studies that examined either perception of emotion in musical fragments (*e.g.,* Gabrielsson & Juslin, 1996) or the fluctuations in tension within short musical pieces (Krumhansl, 1996).

3. Method

The method section is divided into four parts. The three songs and the recordings used in the study are first introduced. This is followed by an explanation of the emotional coding of the songs; and then by a structural coding of the songs. Finally, the measurement method for the performance variables is outlined.

In brief, the method adopted is as follows. The emotion of passages of the songs is characterized in terms of emotional activity, which is high (3), medium high (2) or low (1), and in terms of emotional valence, which is positive (1) or negative (0). Within these passages, tension is low (0) or reaches a local highpoint (1). The structure of the music is coded, including an encoding of the melodic contour (pitch height in semitones), the rhythmic density (number of notes per bar), phrasing (phrase ends are determined), and melodic highpoints. All coded descriptions are made at the level of the bar. The measurements of vocal expression include measurements of the duration of each bar, the average sound level per bar (referred to as amplitude), the rate and extent of vibrato of any long notes within each bar, and the number of pitch glides up and down per bar.

Multiple regression analysis is used to relate the measured variations in performance variables to the coded descriptions of the emotion and structure of the music. In a multiple regression analysis, multiple independent variables are fitted to one dependent variable to account for variations in it. (See the equation for bar duration.) These fits were made for each performance variable (local duration, amplitude, vibrato rate and extent, pitch glides up and down) of each performance of each piece separately.

$$Dur_{bar} = \gamma + \alpha \cdot contour + \beta \cdot highp + \chi \cdot rhythm$$
$$+\delta \cdot phrase + \varepsilon \cdot valence + \phi \cdot activity + \varphi \cdot tension$$

Readers who are not interested in the details of the analyses may skip the method section and continue with the results section.

3.1. Musical material

Three songs by Franz Schubert were used in the study: *Die junge Nonne* (F minor; D828; Op. 43 No. 1; from 1824 or 1825), *Du bist die Ruh*, (E flat major; D776; Op. 59, No. 3; from 1823) and *Gretchen am Spinnrade* (D minor; D118; Op. 2, from 1814). These are abbreviated as JN for *Die Junge Nonne*, DbdR for *Du bist die Ruh* and GaSp for *Gretchen am Spinnrade*. Texts of the songs are given in Appendix A, B and C.

These three songs were chosen because they were regularly recorded throughout the twentieth century, especially DbdR and GaSp, and contain a sequence of different emotions, especially JN (see below). In addition, DbdR and GaSp are contrasting in overall emotion: DbdR is positive and calm overall, while GaSp is negative and restless.

The text of JN, written by Jacob Nicolaus Craigher in 1823, is about a young nun whose heart has been tormented, like the storm raging outside. She expects to find rest and peace within the convent in an eternal marriage to God. Her desire to become a nun is a desire to escape from the torments of earthly life through death (Reed, 1997). Schubert's musical setting exaggerates the emotions suggested by the text to such an extent as to add strong undertones of mental disturbance and sexual fantasy (Leech-Wilkinson, 2007); for more details, see Leech-Wilkinson (this issue).

The text of DbdR is a poem by Friedrich Rückert, written in 1819-1820, that seems to address someone, or perhaps a concept or state of mind, whom the protagonist loves. The beloved gives peace and comfort, and the protagonist wishes her or him to return. According to Reed (1997), the text is an expression of religious devotion, where the longed for love is as much a longing for an ideal world as for a beloved person.

The text of GaSp, by Johann Wolfgang Goethe, is a passage from *Faust* written in 1774-1775. Gretchen is in her room sitting at a spinning wheel. She expresses her passion and mourns her unsatisfied love (for more details, see *e.g.,* Reed, 1997). The famous piano accompaniment of the song refers to the movement of the spinning wheel and illustrates Gretchen's restlessness. The spinning wheel stops for only one moment in the song, where Gretchen is thinking about Faust's kiss. This culmination point of arousal is accompanied by a dominant seventh chord that becomes a diminished seventh and is not resolved until the return to the opening theme a couple of measures later. The sexual fantasy will not lead to Gretchen's happiness: instead she will die in Faust's kiss.

Eight recordings of each song were collected; some can be found on the CHARM discography website.[1] For the recordings up to the 1940s, inclusion was mostly dependent on availability. New CD releases of old recordings were used as well as some custom-transferred recordings from original 78s. Where necessary, clicks were filtered out, because of their effect on measurements of amplitude. All recordings are of female singers, except for DbdR, which is regularly performed by male singers as well: a recording of DbdR by Dietrich Fischer-Dieskau, from 1969, was included. A list of recordings is given in Appendix D. Reference to these recordings is made using initials to refer to the singer and a two-digit number to refer to the year of recording.

The result is a partly systematic collection of recordings spanning the period from 1907 to 1977 with an emphasis on the period between 1920 and 1950. The collection includes famous singers from different generations, in some cases performing more than one of these songs. Most of the singers are very well known for their interpretations of Schubert songs, and also for their operatic performances, except Kathleen Ferrier and Elena Gerhardt who mostly focussed on concert work. Short biographies of these singers can be found in the New Grove Dictionary of Music and Musicians (2001, 2nd edition) and in Kutsch & Riemans (1987).

3.2. Emotional coding

In the analyses of the performances, the focus is on vocal expression; therefore only the bars with voice were included in the analyses.

To create a workable methodology, emotion was defined as active (2), medium active (1) or calm (0); valence as either positive (1) or negative (0); and tension as either high (1) or low (0). The valence and activity of successive bars in a passage or phrase were assumed to be the same, so the change in emotional character changes approximately every 8 bars, with each passage generally containing one emotional highpoint.

In JN, the distinction between passages with high and low activity was made primarily on the basis of melodic and rhythmic characteristics. Active passages with a triadic melody and a melodic range larger than a fourth alternate with calm passages with longer notes and stepwise melodic movement. This distinction between passages continues throughout the piece and is accompanied by a related change in the text (see Appendix A).

The distinction between passages with positive and negative valence was made on the basis of the text and harmony. JN starts negative in valence, with its reference to a storm, restlessness and the grave; in bar 52, the mood becomes more positive, when the possibility of salvation is introduced. Chromatic and enharmonic harmonies from bars 63 to 71 reintroduce an element of negative valence, representing disturbance as the nun calls on her bridegroom for salvation: she begs for salvation. The mood turns more consistently positive from bar 71 till the end, when the text speaks about the peaceful cloister bell that leads the nun to eternal heights and marriage to her Saviour. Here major mode prevails.[2]

Tensional highpoints were defined on the basis of rhythmic and melodic accents. A recurring feature is a melodic ascending leap that accentuates, for example, *heulende Sturm* and *zittert das Haus*. Another recurring feature is rhythmic intensification: *Wie das Grab* enlarges the slowing of rhythm that was already present in the preceding *Und finster die Nacht*. In the second half of the song, tensional and melodic highpoints coincide around the middle or in the second half of a melodic line: this is consistent with Krumhansl (1996), who showed that tension contours tend to increase after section beginnings and decrease at section endings.

DbdR is ambiguous in emotion. It is in the major mode and is generally positive, but not completely so: the love is as yet unfulfilled. The ambiguity is apparent from the melody that generally moves upwards, but its upward motion alternates with downward motion to a lower register. As a simple description of melodic valence, the characterization of passages as positive or negative follows this alternation between prevailing upwards or downwards melodic motion (see Appendix B for a full description).

The activity in DbdR is generally low. The melody and rhythm intensify in the course of the music; the second stanza (bar 16-25) is more intense than the first in rhythm and text. The first halves of the fifth and final stanzas (bar 54-61 and bar 67-76) have the strongest intensity, with a crescendo, harmonic modulations and large melodic range (here activity is highest). In the second half of these stanzas, activity drops again to the level of the second and fourth stanza (activity is intermediate).

Within each sub-phrase (an eight to twelve syllable couplet in DbdR), one bar is a melodic focal point (high in tension). Highpoints in tension are accented in melody and/or rhythm: for example, *Friede* and *was* are melodic highpoints, while the bars containing *Lust* and *Aug'* are also rhythmically differentiated (more active). The following tensional highpoints are similarly either the endpoints of upwards melodic motion, or differentiated rhythmically and relatively high.

GaSp is overall a negatively valenced song: Gretchen's peace is gone and she will never find it again without Faust's presence. Within this overall mood, a turn towards a more positive state can be seen in bars 51 to 68 (first note only), where Gretchen sums up the attractive characteristics of her lover (see Appendix C). This relatively positive valence is also present from bars 85 to 92 , reducing to 96, where she thinks about his touch and kiss. However this positive longing has turned negative by bar 97, where the implication of dying in his kiss (anticipated at bar 68, second note) becomes clear.[3]

In GaSp, clear indications are given in the score of growth of activity: crescendos and higher pitch ranges indicate more active passages. Activity is highest in passages containing climactic moments indicated by sforzati (bar 63 and onwards) and fortissimo (e.g. bar 93 and onwards). All returns to the opening stanza are interpreted as returns to a state of lower emotional activity.

As with the other two songs, tension builds up and reaches a local highpoint within a span of four to eight bars. These focal points are often high in melodic pitch, for example <u>*Herz*</u> in bar 5 and <u>*nimmer*</u> (from <u>*nimmermehr*</u>) in bar 10, and accented in harmony (almost all have a dominant function and often introduce a modulation). Examples include bar 68, the moment of Faust's kiss, and bar 111, the dramatic climax at the end (see also Appendix C).

3.3. Structural coding

As mentioned before, the structural descriptions of the songs include melodic contour, melodic highpoint, rhythmic density, and phrase structure. These are simple descriptions of the musical structure that can be encoded in a straightforward manner. Like the emotional characterizations, the structural characterizations were made at the level of the bar.

<u>*Melodic contour*</u> is the average pitch for each bar of the melody in semitones counted from middle C. For example, in GaSp the first bar with voice consists of a half note with pitch A, followed by an eighth note with pitch B flat, followed by an eighth note with pitch A. The average pitch of the melody of this bar is $\frac{(4*9)+10+9}{6} = 9.17$.

<u>*Melodic highpoint*</u> distinguishes bars with absolute highest pitches. Melodic highpoints are high both globally and locally, which means that they are high in absolute terms as well as relative to the immediate musical context.

Rhythmic density gives the number of melody notes per bar. If short notes prevail within a bar, more notes are present than if longer notes prevail.

Phrase structure distinguishes phrase endings (1) from the rest of the phrase (0). Phrases generally follow the structure of the text. DbdR has a phrase ending after every sentence (after *mild*, *stillt* and *schmerz* etc, see Appendix B). GaSp has a phrase ending after every second sentence (after *mehr*, *vergällt*, and *zerstückt* etc, see Appendix C). JN has a phrase ending after every sentence (after *Sturm*, *Haus*, and *Blitz* etc, see Appendix A). However, sometimes sentences of the poem are connected and integrated within one musical phrase: this is for example the case for the sentences *Es brauste das Leben* to *wie jetzo der Blitz*.

For GaSp, *fermata* was added as a structural component. This is used only to distinguish bar 68 (which refers to Faust's kiss) from other bars.

Different scores of these songs show slightly different versions, but the interpretations given here are mostly at a sufficiently abstract level not to be affected by these differences. The exception is melodic contour for DbdR, where some singers use a version with different melody (and harmony) in bars 56 and 70. This deviation has not been included in the analyses, because it concerns only two out of 59 vocal bars.

3.4. Measuring expression in recorded performances

Measurements of expression were made at the bar level. This facilitates comparability between different performance measures, comparability between performances, and comparability with the analysed characteristics of the music per bar. Moreover, choosing the bar level instead of a smaller unit has the benefit that variations due to inaccuracy in measurements are relatively small.

All measurements were made using PRAAT.[4] The location of bar onsets was determined manually using an annotation tool. By zooming in to a window size of short duration (around 200 ms), measurement error can be reduced to about five ms. However the actual measurement variation is higher, because the onset of vocal tones is difficult to determine exactly and consistently. As a rule, the onset of a vocal tone coincides with the onset of the vowel, which is also the onset of the periodic part of the signal (for a similar methodology see Ashley, 2002).

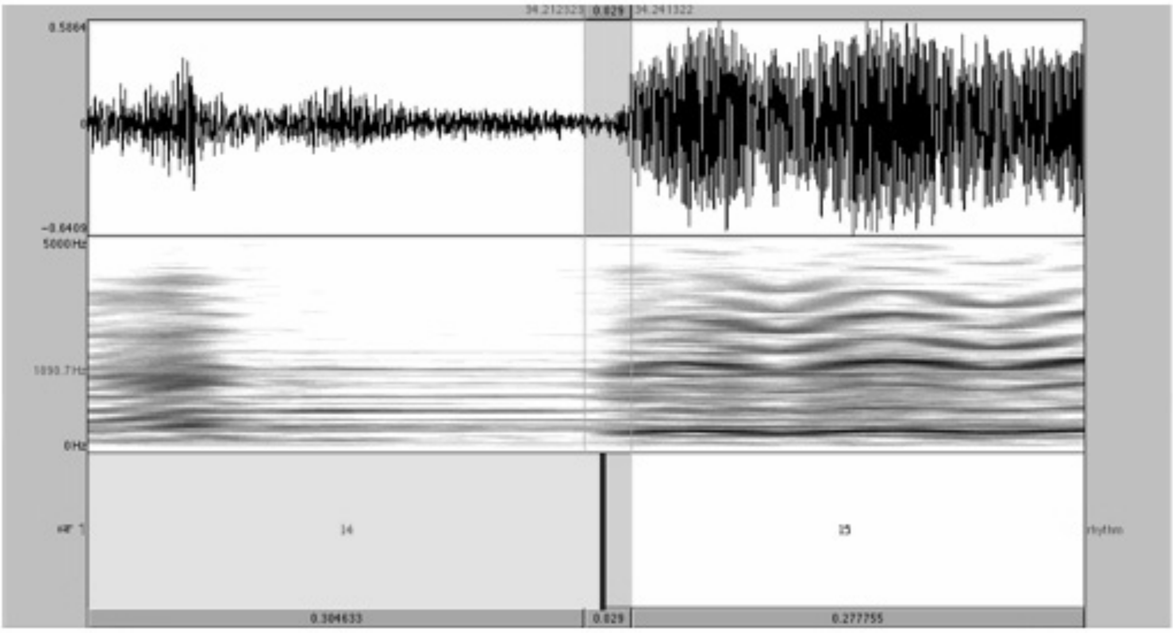

Figure 1: Annotation of sound file with bar-lines. Dotted lines indicate the estimated accuracy of the manually located bar onset.

Figure 1 illustrates an estimation of the maximum inaccuracy that may arise in measuring the onset of a vocal tone. The beginning of the tone is clearly visible. However, for other tones, the onset might be partly masked. Lines are drawn at repeated measurements that indicate a likely range within which the onset would be found. In this case the range is 29 ms. The average bar durations of the songs are 3.3 seconds for JN, 2.9 seconds for DbdR and 1.7 seconds for GaSp; 29 ms is therefore less than 2% of the smallest average bar duration. The smallest measured standard deviation of bar IOI for individual performances is 70 ms (Elly Ameling singing JN). The maximum inaccuracy is thus less than half the minimum measured standard deviation. Importantly, accuracy tends to increase with recordings of modern performances that show smaller tempo variations, because of the lack of background noise in modern recordings and the strong alignment of the piano and voice. Measurements of early recordings tend to be less accurate, but this is compensated by the larger tempo variations. Moreover, the regression analysis distinguishes signal from noise by highlighting only systematic relationships between structural and emotional characteristics of the music and measured performance variations.

After bar onsets had been located, time information was used to calculate the average amplitude of the sound signal per bar in dB above the hearing threshold. This is an average of the complete signal within a time window including piano, voice and any recording artefacts. As mentioned before, reproductions were used without clicks, but hiss is present in older recordings: since it generally has stable characteristics, however, hiss only affects measurements of global amplitude and not variations in amplitude.

To measure vibrato, a relatively high and long note was chosen as representative of each bar. The vibrato rate is the duration of the part of the note that contains vibrato divided by the number of vibrato cycles: it was measured between two vibrato peaks or troughs (see Figure 2). For vibrato extent, one large vibrato cycle was selected and the maximum and minimum values within that selection measured in semitones[5]. Choosing long notes and large cycles for measurements of vibrato rate and extent respectively facilitates measurement accuracy.

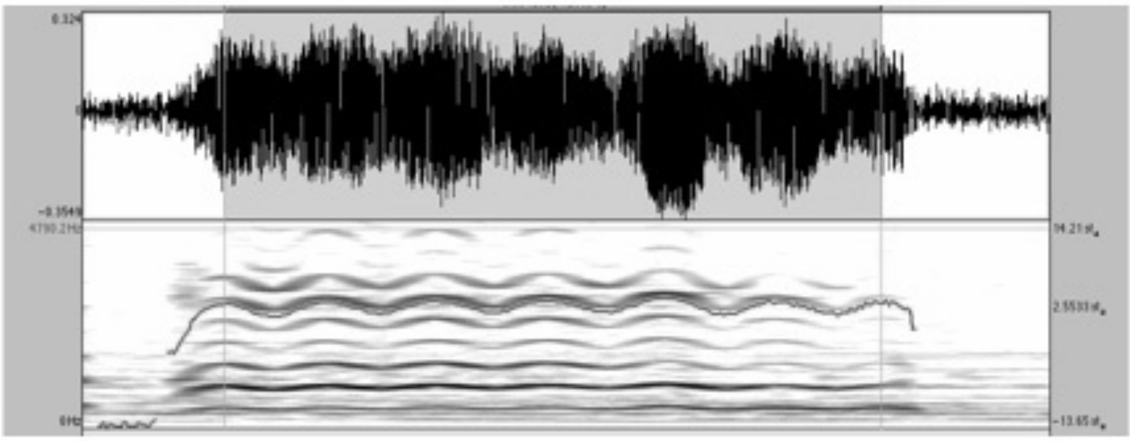

Figure 2: Analysis of pitch within a sound file. Dotted lines indicate the analysis of the time distance between two vibrato peaks.

The detection of pitch cannot deal with polyphonic signals, but it does detect periodicities within a noisy background. Therefore as long as the voice signal is considerably louder than the accompaniment, PRAAT detects pitch accurately, particularly in the case of female voices. Pitch detection is generally better in older recordings, where the voice is much louder than the accompaniment, than in modern recordings (from the 1960s onwards), where piano and voice are in equal balance: in these modern recordings, only parts of sung tones, often the louder parts, are correctly detected and can be used to measure fundamental pitch. Because only one larger vibrato cycle per note per bar was needed for the measurement of vibrato extent, this was not a real problem.

Finally, measurements were made of pitch glides up and down per bar. This was done by ear with assistance of visual information. A distinction was made between prominent pitch glides and less prominent pitch glides: prominent pitch glides are relatively slow and loud, while less prominent pitch glides are softer. In the present analyses, these distinctions are collapsed into two categories of glides, up and down, but in future analyses it may be important to distinguish between the types.

4. Results

The results are split in three parts, the first of which focuses on the overall characteristics of the performances, the second on the similarity in expression between performances, and the third on the relationships between expression, structure and emotion.

4.1. Performance characteristics

To get an impression of the overall characteristics of the performances, Figures 3 to 7 plot the average of the measured performance variables and the standard errors. Standard errors are a measure of the variation around the mean, and in the third section an attempt is made to explain this variation. For now, the standard error bars are interpreted as an indication of the amount of variation: for instance, in the case of bar duration this is taken as a measure of the amount of tempo rubato within a performance.

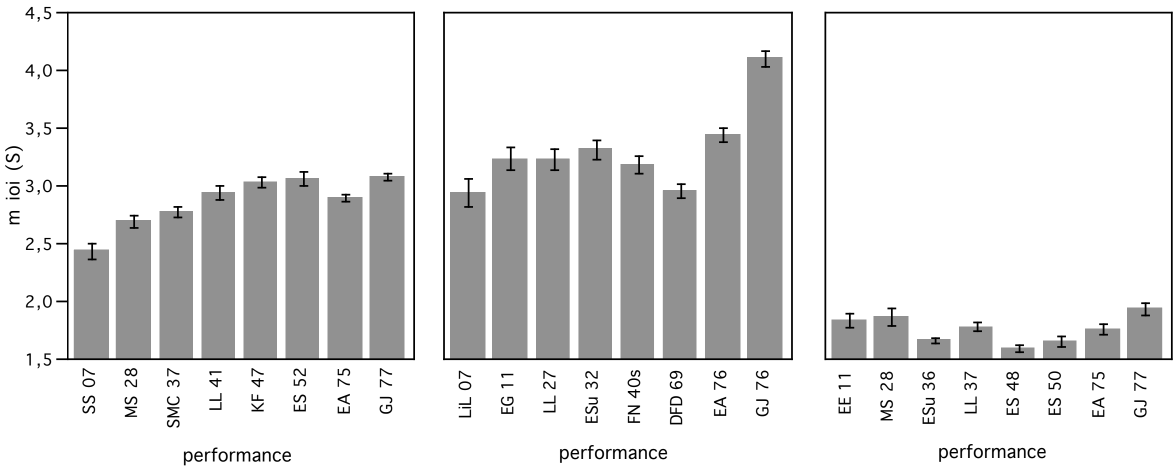

Figure3: Average bar duration and standard errors for each performance of each song.

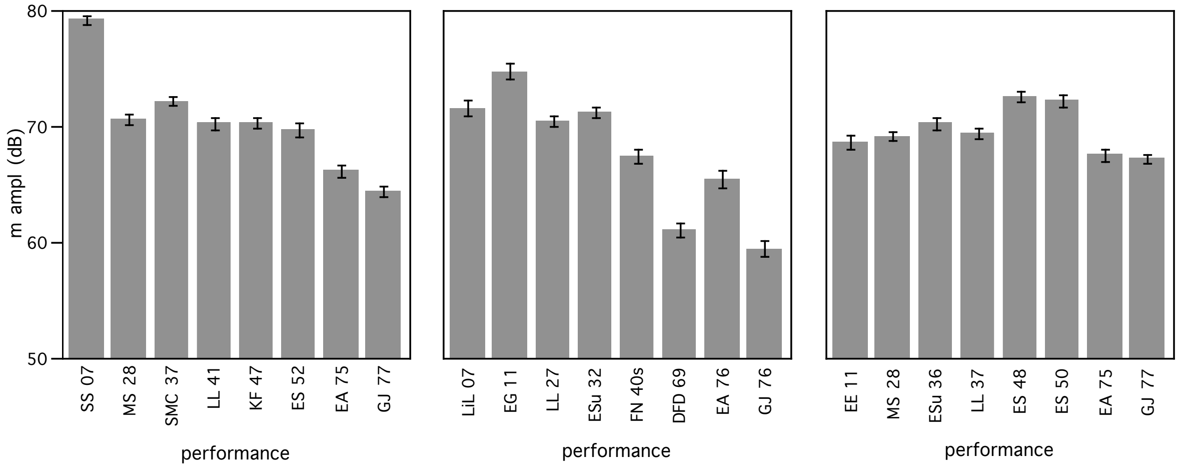

Figure 4: Average bar amplitude and standard errors for each performance of each song.

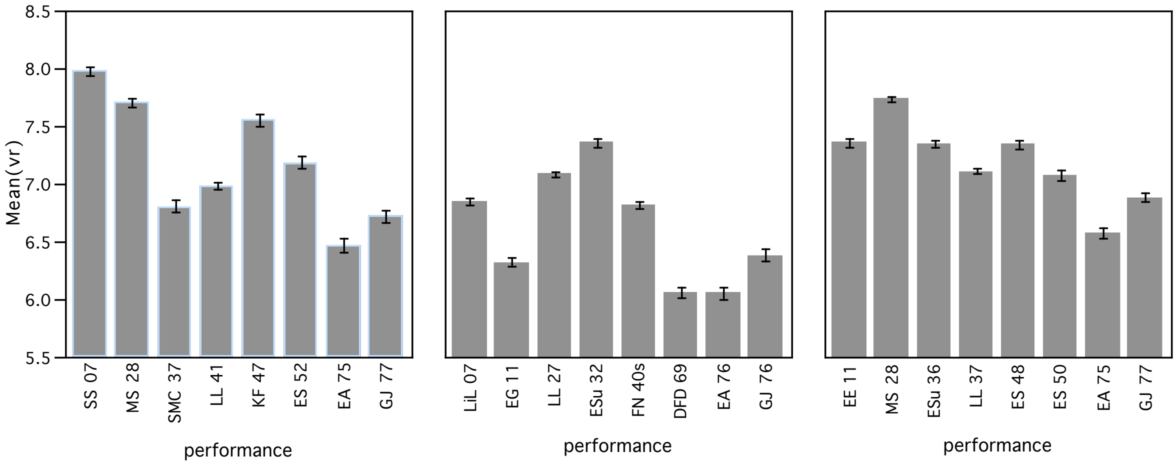

Figure 5: Average vibrato rate and standard errors for each performance of each song.

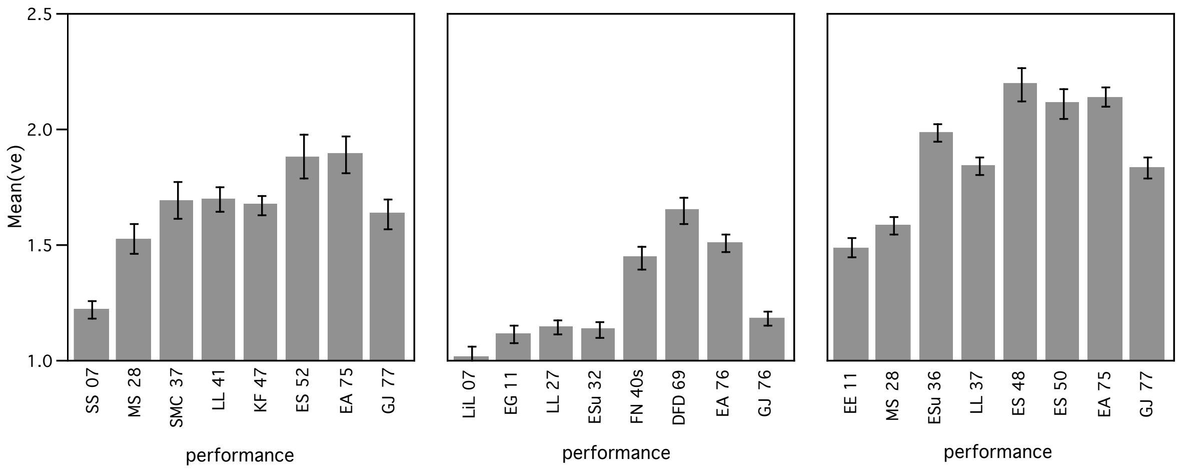

Figure 6: Average vibrato extent and standard errors for each performance of each song.

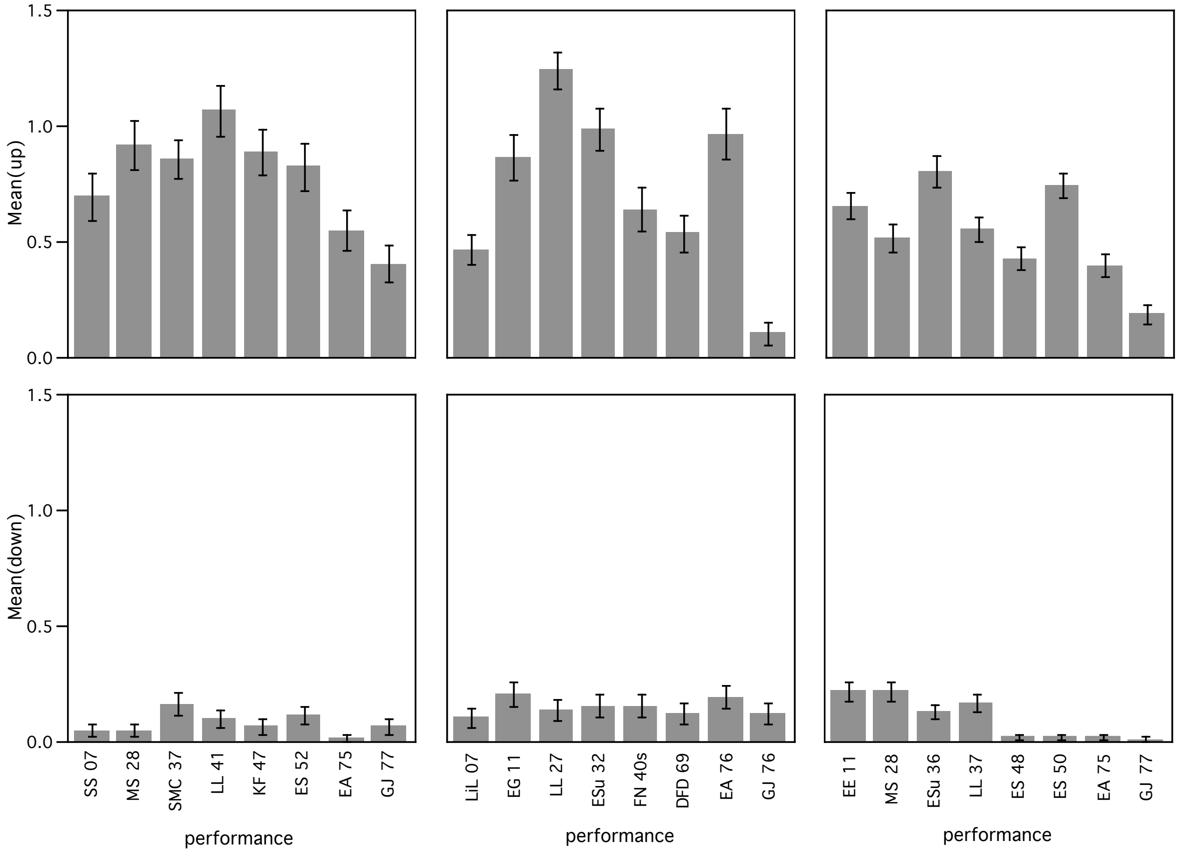

Figure 7: Average number per bar and standard errors of upward (top) and downward pitch glides (bottom) for each performance of each song.

The figures show clear differences in performance characteristics between pieces, performers and date of recording. Differences between pieces tend to correspond with the character of the songs: GaSp is performed fastest, with the highest average vibrato rate and extent, while DbdR is performed slowest, with the lowest average vibrato rate and extent. These differences correspond with the "restlessness" of GaSp and the relative calmness of DbdR. Additionally, GaSp tends to have fewer upward glides than the other two pieces, while the early performances of GaSp tend to have more downward glides. These tendencies correspond with the overall negative character of GaSp.

Systematic changes over time can be observed for all performance characteristics in Figures 3-7: variation in bar duration, overall amplitude, and vibrato rate tend to decrease over time, while vibrato extent tends to increase. The number of upward pitch glides was highest for the recordings from the mid-twentieth century and lowest for the end of the century. The number of downward pitch glides was generally small, especially for performances recorded in the second half of the twentieth century.

These tendencies over time suggest considerable changes in performing styles. However the measurements of amplitude should be interpreted with caution: the high sound level for early recordings may partly be due to the constant presence of background noise from the recording. Additionally, performers in early recording sessions were expected to maintain a high amplitude level in order for all notes to register on the recording, and may not be representative of practice outside the studio.

4.2. Similarity in expression

The previous section showed considerable differences in overall performance characteristics for performances from different periods. This section compares the expression of different performances across periods. Table 1 shows the average correlations between performance variations for each variable and each song: the correlations between variations in bar duration (IOI) and in amplitude are generally high ($r > .55$). The correlations between variations in vibrato are high for JN, but low for the other two songs. Finally, the correlations between the positions of upward pitch glides are generally low, while the correlations between the positions of downward pitch glides are high for DbdR and JN, but low for GaSp.

Table 1: Average pair-wise correlation between measured performance variables per song.

Song	IOI	Ampl	V Rate	V Ext	Up	Down
JN	.64	.60	.44	.50	.22	.44
DbdR	.67	.62	.17	.28	.27	.71
GaSp	.55*	.75	.24	.19	.16	.23

* Correlation is calculated for all bars excluding the fermata. The fermata would artificially increase all correlations.

To test whether correlations are smaller for performances recorded wider apart in time, the relationship between the time-difference and pair-wise correlations between two performances was tested using a regression analysis with time-difference as independent variable and correlation as dependent variable. This relationship was significant for amplitude and vibrato extent in JN (F (1, 26) = 12, $p < .01$; F (1, 26) = 5.7, $p < .05$), for amplitude in DbdR (F (1, 26) = 4.9, $p < .05$), and for bar duration, vibrato rate and position of upward pitch glides in GaSp (F (1, 26) = 6.6, $p < .05$; F (1, 26) = 6.6, $p < .05$; F (1, 26) = 15, $p < .001$).

In sum, the association between performance variations was rather high for amplitude and bar duration, and, for some songs, for vibrato and pitch glides. This association was often irrespective of the time difference between two performances. However, in some cases, the association between performances tended to decrease with increasing time-difference between recording dates, notably for GaSp, providing some evidence for a changing tradition in interpretation.

4.3. Relationships between expression, emotion and structure

To examine performers' strategies for communicating the emotion and structure, the relationship between variations in performance variables, coded emotion and structure was tested using multiple regression analysis. This was done for each performance of each piece separately. Table 2 presents the results for one performance of JN: it shows the explained variance of the model for each performance variable, as well as the direction and the significance levels of significant relationships. This specific performance, by Lotte Lehman, was chosen as an example, because the model accounts well for variations in all the performance variables of her performance. Her performance is representative of the kinds of significant relationships that were found.

Table 2: Results of the fit of the multiple regression model to the performance of JN by Lotte Lehman: Explained variance, $\underline{p}$ value (4 levels), and direction (+ or -) are given for the fit of the model to each performance variable.

Fit	IOI	Ampl	V Rate	V Ext	Up	Down
LL 41	0.50	0.57	0.66	0.63	0.33	0.28
Contour		**** (+)				** (-)
Rhythm	** (+)	** (+)	*** (+)			
Highpoint					* (-)	** (+)
Phrasing						
Valence	**** (+)	** (-)	* (-)	** (-)	** (-)	
Activity	*** (-)	** (+)	*** (+)	**** (+)		
Tension						

* $\underline{p} < .05$, ** $\underline{p} < .01$, *** $\underline{p} < .001$, **** $\underline{p} < .0001$

Table 3: Summary of the results of the fit of the multiple regression model. The number of performances that showed a significant relationship between a musical characteristic (rows) and a performance variable (columns) are reported as well as the minimum, average and maximum explained variance ($\underline{R^2}$ at the top of each cell).

Fit	IOI	Ampl	V Rate	V Ext	Up	Down
JN						
Min/Ave/Max $\underline{R^2}$	.33 .50 .62	.35 .56 .74	.34 .49 .69	.21 .59 .77	.18 .30 .47	.12 .22 .43
Contour		7+	1-			2-
Rhythm	2+	5+ 1-	6+	4+	3+	
Highpoint		1+	2+		1-	2+
Phrasing	5-	1-		1+		
Valence	8+	3-	3-	6-	2-	1+
Activity	5-	5+	3+	7+	3+	1-
Tension	7+	3+		1-		
DbdR						
Min/Ave/Max $\underline{R^2}$	.41 .61 .80	.52 .68 .79	.22 .47 .73	.24 .44 .64	.16 .38 .51	.47 .53 .61
Contour		7+	2- 1+	1-	3-	8-
Rhythm	8+	1+	4+	1+	1+	8-
Highpoint	2+	4- 1+	1+ 1-			8+
Phrasing		4-	2+ 1-	3-	5-	3-
Valence	7-	1+ 1-	1+	3+	4+	8-
Activity	1-	4+	2+	4+	3+	8+
Tension		1+		2-		
GaSp						
Min/Ave/Max $\underline{R^2}$	.30 .72 .94	.53 .66 .81	.04 .16 .23	.13 .21 .31	.04 .11 .20	.02 .17 .36
Contour		8+		2-		
Rhythm	3-	2+	1-	1+		1+
Highpoint		1+	1+ 1-	2-		1-
Phrasing	1+	1+		1-		
Valence	6-	5+	2+	1+		
Activity	4-	4+		3+	1-	1+ 1-
Tension		2+		1+		2+

Table 3 gives a summary of the fit of the model for all performers. It shows the minimum, average and maximum explained variances as well as the number of performers who showed a significant relationship in a positive or negative direction. To interpret this direction, it is important to know that a low value of the independent variable corresponds to low pitch, low rhythmic density, not a local highpoint in pitch, not a phrase end, negative valence, low activity and low tension, while a high value corresponds to the opposite. In addition to the reported results, fermata was significantly related to bar duration for all performances of GaSp.

Considering each performance variable in turn, Table 3 shows that, for bar duration, valence is most often significant. The direction of this relationship is positive for JN (longer duration for positive passages) and negative for DbdR and GaSp. The other effects that account for variance in bar duration in JN are tension, activity, phrasing and rhythm: bar duration increases at moments of high tension, decreases with activity (tempo increases), decreases at phrase ends, and increases in bars with higher rhythmic density. The decrease in bar duration at phrase ends is unexpected: a lengthening at phrase ends is more common. This faster tempo may be related to the specific phrase ends in JN, which often coincide with the beginnings of the following phrases.

In DbdR, the relationship with rhythmic density is also strong. As in JN, but in contrast to GaSp, bars with high rhythmic density are lengthened. Additionally, bar duration increases at highpoints, which is related to the lengthening of bars at moments of high tension in JN.

In GaSp, most variation in bar duration is due to the fermata. Relationships with valence and activity are also significant, and, for some singers, with rhythm and phrasing. These latter effects differ in direction from the other songs. The effect of phrasing is, as expected, a lengthening of bar duration, while the effect of rhythm is a shortening of bar duration at moments with higher density. This may be related to the specific rhythmic characteristics of GaSp: intense passages in GaSp have relatively long notes and therefore fewer notes within a bar (such as the moment of Faust's kiss).

For the variation in amplitude, the effect of contour and activity is most consistent: in almost all performances amplitude increases with contour, and in many of the performances with activity. For the one performance in which contour is not significant, highpoint is significant: amplitude is higher at pitch highpoints. All other effects also often reach significance. In JN and GaSp, the effect of rhythm is an increase in amplitude in passages with more notes. The effect of phrasing is mostly negative: amplitude is softer at phrase ends. The effect of valence is negative in JN, but positive in GaSp: in JN, positive passages are accompanied by softer dynamics, while in GaSp positive passages are louder. Finally, dynamics tends to increase with tension, for some singers in each song.

Although vibrato may be expected to vary less systematically than tempo or amplitude, remarkably high explained-variances are obtained for most performances of JN and DbdR. The effect that accounts for most of the variation in vibrato rate is rhythm: vibrato rate is faster when the note durations are relatively short. Additionally, vibrato rate and valence are related negatively for JN and positively for GaSp, analogous to the other relationships with valence for these songs. Vibrato rate is positively related with activity for some singers in JN and DbdR.

The strongest relationships for vibrato extent are with valence and activity. The relationship with valence depends on the song (negative for JN, positive for the other two songs). The relationship with activity is consistently positive. Additionally, all three songs show a positive relationship between vibrato extent and rhythm for some singers. Other, occasional, relationships include relationships with phrase end, contour and highpoint. These are generally negative.

The position of upward and downward glides is less consistently related to the variables of the model, with the exception of downward glides in DbdR. Nevertheless, some tendencies are clear: increases in rhythm and activity correlate with increases in the number of upward pitch glides in JN and DbdR, and downward pitch glides in DbdR. On the other hand, the number of pitch glides decreases in JN and DbdR with increasing melodic contour and at phrase endings. Relationships between pitch glides and valence again depend on the song. Additionally, more downward pitch glides are performed at highpoints in DbdR.

Taking these results together, it is clear that the direction of significant relationships is generally consistent across different performances of the same piece: significant relationships show opposite directions in a few instances only. However, the direction of significant relationships may differ across pieces, even for the stronger relationships. Additionally, relationships with valence and activity are most often significant, followed by relationships with rhythm. Effects of highpoint, phrasing and tension are less often significant. Contour is especially significantly related to amplitude, but also to downward pitch glides in DbdR.

These summary results do not show changes in strategy over time: instead they suggest considerable consistency over time, given that every performance variable shows at least one relationship that all or most of the singers share. Nevertheless, it is possible that certain relationships occur more often in one period than in another. To investigate this possibility, the eight performances of a song were grouped by pairs. For each pair shared significant relationships were counted. Next, the overlap with performances of other groups was established by counting the number of relationships from these significant relationships that were also shown by at least one performance of another group. For example, the two performances of Gr1 for JN (SS 07 and MS 28) share three relationships; two of these are also present in performances of Gr2, and all three relationships are present in performances of Gr3 and Gr4 (see first row of Table 4).

Table 4: Results of a comparison between shared significant relationships within and between groups of two performances. Performances are grouped according to recording date: Group 1 (Gr1) consists of the two earliest performances. Gr2 consists of the third and fourth performance in time, etcetera. Shared relationships within a group are first established (diagonal). Next, overlap with other groups is counted for these relationships (rows).

Fit	Gr1	Gr2	Gr3	Gr4
JN				
Gr1	3	2	3	3
Gr2	1	8	7	4
Gr3	7	9	10	9
Gr4	7	8	9	9
DbdR				
Gr1	8	8	6	8
Gr2	10	10	9	9
Gr3	10	9	10	9
Gr4	10	10	9	12
GaSp				
Gr1	2	2	1	2
Gr2	2	2	2	2
Gr3	3	3	5	3
Gr4	3	2	2	3

If relationships are time-dependent, a decrease in overlap between groups that are further apart in time should show up. Table 4 shows that this was not generally the case: the comparisons with Gr2 of JN form the only exception. The overlap between significant shared relationships of Gr2 and significant relationships in performances of Gr1 and G4 is relatively small. Despite this exception, Table 4 shows that the shared significant relationships within a group are generally also relationships shared with performances from other groups.

As a final analysis of possible changes over time, the association between recording date and the explained variance of the model fit was tested using a regression analysis with date as independent variable and explained variance as dependent variable. This yielded a significant result for amplitude in JN (F (1, 6) = 12, p <.05), vibrato rate in DbdR (F (1, 6) = 69, p <.001), and bar duration in GaSp (F (1, 6) = 9.8, p <.05). For all other variables of the performances of these songs, the explained variance of the model fit was independent of recording date.

5. Discussion

The main findings from this study can be summarized as follows. First of all, changes in performing style over time were observed that included changes in every measured performance variable: local duration, amplitude, vibrato and pitch glides. Secondly, in contrast to these changes in performing style, strong similarities in vocal expression were observed, especially for local duration and amplitude, but also partly for vibrato and pitch glides. Although there was a tendency for correlations between performances to decrease with increasing distance between recording dates, overall the correlations between performance variations were rather high. Thirdly, this similarity in vocal expression was even more apparent from the analysis of the performing strategies used to express the emotion and structure of the songs: the stronger relationships between expression, emotion and structure were shared among performers irrespective of recording date. Fourthly, the model was rather successful in accounting for performance variations. Explained variances were often around 50% or higher not only for bar duration and amplitude, but also for vibrato and sometimes for the position of pitch glides. In particular, the success of the model can be attributed to systematic relationships with the coded emotion of the songs, although structural aspects such as rhythmic density played a significant role as well. For some performance variables, the explained variance increased with recording date, suggesting that these variables were more systematically applied in modern performances than in older performances, or at least that the model accounts better for expression in modern performances than earlier ones. Finally, the relationships between expression, emotion and structure were generally consistent in direction across pieces and performers, except for the relationships with valence and rhythm, which depended on the performed song.

These findings emphasise the generality of performance strategies to communicate the structure and emotion of songs across diverse performing styles. This is evidence for the theory that expression is generated from a representation of musical structure (*e.g.,* Clarke, 1988; Palmer, 1997), while suggesting that emotion is at least as important. However, the study also suggested the complexity of these generative processes by highlighting dependencies on the performed song.

This does not mean that the performers interpreted the songs in exactly the same way. Each song has multiple layers of meaning, and different performers may communicate different interpretations. The procedure that was used in this study highlighted the similarities, while, at a different level of analysis, differences may be much more apparent. For example, the intensity of "negative" or "positive" passages of the different songs may be different for different performers, as may the communicated levels of "relaxation" or "excitement". These hypotheses have been tested for a selection of passages from JN in perceptual experiments (Timmers, 2007), while Leech-Wilkinson (this issue) discusses diverse interpretations of JN at greater length.

As for the dependencies of the relationships on the performed song, in JN, valence related positively with bar duration, and negatively with amplitude, vibrato rate and vibrato extent. In DbdR and GaSp, the relationships with valence were the other way around. A possible explanation is that in JN the positive passages primarily form a resolution of negative tension, and in GaSp the positive passages form an escape from the negative background; while in DbdR the positive passages form the main background, alternating with negative reflections. These differences in relationships between positive and negative passages may explain the observed differences in valence: if positive passages form an escape, positive effects on tempo, amplitude and vibrato may be expected, while if positive passages form a resolution, the expected effect is one of relaxation. Notably, these relationships are in addition to those relating to emotional activity.

Relationships with rhythm also varied with musical piece. For JN and DbdR, bar duration, amplitude and vibrato increased with rhythmic density - as might be expected. For GaSp, the relationships with rhythm were significant only for a few performers, while the relationship with tempo was opposite from the other songs: tempo was faster in bars with fewer notes. GaSp does not have strong variation in rhythmic density: the number of notes per bar tends to be three and sometimes two, while there are bars with only one note at intense moments (*e.g.,* Faust's kiss). The smaller variation in rhythmic density makes the observed relationships less reliable, while the specific location of the bars with few notes and the lack of bars with many notes may explain the decrease in tempo for bars with fewer notes.

These discussions of the observed relationships highlight some of the difficulties in applying models to performances of complex musical pieces. One difficulty is that structural and emotional characterizations of different pieces may not be directly comparable: for example, variations in rhythmic density in one piece may not be exactly comparable to those in another. Another difficulty is that in complex musical pieces different aspects of the music tend to correlate: some observed effects might be partly due to other variables that are not taken into account, or they may actually not be significant but merely correlate with other variables that are. For example, tension, highpoint and contour correlate, but it is unlikely that all three are independently significant; amplitude tends to increase with pitch contour, but the highest points are often softer.

The findings and conclusions of this study are suggestive rather than definitive. There is no hard evidence that the singers conceived the music in the way proposed here. The model simply measures the extent to which the performances agree with or contradict its assumptions: its strength is in testing predictions and summarising trends in a collection of performances. Additionally, the analyses emphasized the similarities between performances by focussing on the bar-level and by focussing on systematic relationships with predefined structural and emotional aspects of the music: they therefore de-emphasise smaller scale differences below the bar level, as well as idiosyncratic interpretations of the music.

This study could be extended in a number of respects. It would be of interest to add perceptual and qualitative data to the quantitative analyses presented here; the interpretation of emotions in the three songs and the influence of a performer on this interpretation of emotions could be tested using perceptual experiments (see Timmers, 2007). The study could also be elaborated to include performing strategies related to smaller scale structures at the note level or more complex structures that go beyond the surface level. It could also be complemented by a more controlled investigation of expression in these songs, using present-day performers.

Despite its limitations, however, the study has presented a wealth of information that is relevant in its details as well as in its overall results. Relevant details concern, for example, the relative absence of significant relationships with phrase structure, the local lengthening of moments of high tension, and the strong relationship between melodic contour and amplitude, both confirming and disconfirming previous observations for instrumental performances (e.g. Todd, 1985, 1992; Schubert, 2001; Sundberg *et al.*, 1989). More generally, it can be concluded that the data presented in this paper highlighted clear similarities in the performing strategies adopted by different singers to express structural and emotional characteristics in three Schubert songs. These similarities occurred despite equally clear differences in performing styles. The similarity in expressive variations and their relationships to structural and emotional characteristics of the music give insight into possible generative processes underlying expressive vocal performance. Emotional characterizations of music play as important a role in the generation of expression as do structural characterizations, and performing strategies vary from song to song, providing insight into the operation of such strategies in complex musical contexts.

Acknowledgements

This work was funded by the UK Arts and Humanities Research Council Research Centre for the History and Analysis of Recorded Music (CHARM). It forms part of a wider project on "Expressive gesture and style in Schubert song performance" led by Daniel Leech-Wilkinson.[6] I would like to thank him for collecting the musical material, and for his support and ideas for the study. I would like to thank Daniel Leech-Wilkinson, Eric Clarke, Nicholas Cook, Wim Timmers and two anonymous reviewers for their helpful suggestions on the text. I would like to thank Roger Beardsley for the transfer of some of the recordings.

[1] Some of the recordings can be found and listened to on the CHARM website at King's College London: www.kcl.ac.uk/charm/ or alternatively at www.kcl.ac.uk/music/ksa/.

[2] Of course, this 'solution' can be seen as unconvincingly positive, especially since Schubert continues the storm music in the piano throughout even this most positive section. Darker interpretations of the text can easily be proposed. However, since the focus is on the prevailing surface interpretations of these three songs, more complex readings may be left as possibilities for future research.

[3] Many interpretations exist of this song, and Faust's kiss is not always interpreted as positive (Reed, 1997). In the interpretation adopted here, this ambiguity is apparent from the sequence of emotions rather than from an ambiguity of simultaneous emotions. However, it should be noted that "positive" in this song is never fully positive, but rather relatively positive or a relative "uplifting" from the overall negative background.

[4] A web-link to PRAAT and explanatory texts can be found at www.fon.hum.uva.nl/praat/.

[5] The calculation of fundamental pitch in PRAAT is based on periodicity detection within a time window. The window size depends on the lowest possible pitch (floor pitch) and is smaller when the floor pitch is relatively high as for female voices and high-pitched singing. A smaller time window increases the accuracy and time resolution for pitch detection. These pitch settings have to be specially set for singing, since the pitch range in singing is different from the pitch range in speech. In the current analyses, a pitch floor of 220 Hz and a pitch ceiling of 10000 Hz were used.

[6] A description of this project may be found at www.charm.rhul.ac.uk/content/projects/schubert.html.

References

Ashley, R. (2002). Do[n't] change a hair for me: The art of jazz rubato. *Music Perception, 19 (3),* 311-33

Brown, C. (1999). *Classical & romantic performance practice 1750-1900.* Oxford: Oxford University Press.

Clarke, E.F. (1988). Generative principles in music performance. In J. A. Sloboda (Ed.), *Generative processes in music. The psychology of performance, improvisation and composition* (pp. 1-26). Oxford: Clarendon Press.

Day, T. (2000). *A century of recorded music: Listening to musical history.* New Haven & London: Yale University Press.

Fabian, D. (2003). *Bach performance practice 1945-1975: A comprehensive review of recordings and literature.* Aldersholt: Ashgate.

Gabrielsson, A. & Juslin, P.N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music, 24,* 68-91.

Gabrielsson, A. (1999). The performance of music. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic Press, pp. 501-602.

Gabrielsson, A. & Juslin, P.N. (2003). Emotional expression in music. In R.J. Davidson, K.R. Scherer, & H.H. Goldsmith, (Eds.). *Handbook of affective sciences* (pp. 503-534). Oxford: Oxford University Press.

Hudson, R. (1994). *Stolen time: the history of tempo rubato*. Oxford: Clarendon Press.

Juslin, P.N., Friberg, A., & Bresin, R. (2002). Toward a computational model of expression in performance: The GERM model. *Musicae Scientiae Special Issue 2001-2002,* 63-122.

Juslin, P.N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129 (5),* 770-814.

Krumhansl, C.L. (1996). A perceptual analysis of Mozart's piano sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception, 13 (2),* 401-432.

Kutsch, K.J., & Riemens, L. (1987). _Grosses Sängerlexikon_. Bern: Francke Verlag.

Langer, S.K. (1963/1942). _Philosophy in a new key. A study in the symbolism of reason, rite and art._ Cambridge, Massachusetts: Harvard University Press. 3rd Edition.

Leech-Wilkinson, D. (2007). _The study of musical performances._ Manuscript in preparation.

Meyer, L.B. (1956). _Emotion and meaning in music._ Chicago: The University of Chicago Press.

Ohgushi, K., & Hattori, M. (1996). Emotional Communication in Performance of Vocal Music – Interaction between Auditory and Visual Information. In _Proceedings of the Fourth International Conference for Music Perception and Cognition_ (pp. 269-274), Montreal, Canada.

Ortony, A., & Turner, T.J. (1990). What is basic about basic emotions? _Psychological Review, 97,_ 315-331.

Palmer, C. (1989). Mapping musical thought to musical performance. _Journal of Experimental Psychology: Human Perception & Performance, 15 (2),_ 331-346.

Palmer, C. (1997). Music Performance. _Annual Review of Psychology, 48,_ 115-138.

Philip, R. (1992). _Early recordings and musical style: Changing tastes in instrumental performance, 1900-1950._ Cambridge: Cambridge University Press.

Reed, J. (1997). _The Schubert song companion._ Manchester: Mandolin. First published in 1985 by Manchester University Press.

Russell, J.A. (1980). A circumplex model of affect. _Journal of Personality and Social Psychology, 39,_ 1161-1178.

Sadie, S., & Tyrrell, J. (2001). _The new grove dictionary of music and musicians._ 2nd Edition. London: MacMillan Publishers.

Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. _Psychological Bulletin, 99,_ 143-165.

Scherer, K.R. (1995). Expression of emotion in voice and music. Journal of Voice, 9, 235-248.

Scherer, K.R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R.J. Davidson, K.R. Scherer, & H.H. Goldsmith (Eds.), Handbook of affective sciences (pp. 433-456). Oxford, New York: Oxford University Press.

Scherer, K.R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? Journal of New Music Research, 33(3), 239-251.

Shaffer, H. (1992). How to interpret music. In M.R. Jones, & S. Holleran (Eds.), Cognitive bases of musical communication (pp. 263-278). Washington: American Psychological Association.

Schubert, E. (2001). Continuous measurement of self-report emotional response to music. In P.N. Juslin, & J.A. Sloboda, (Eds.), Music and emotion: Theory and research (pp. 393-414). Oxford: Oxford University Press.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. Music Perception, 21 (4), 561-585.

Sloboda, J.A. (1983). The communication of musical metre in piano performance. Quarterly Journal of Experimental Psychology, 35 (A), 377-396.

Sloboda, J.A., & Lehmann, A.C. (2001). Tracking performance correlates of changes of perceived intensity of emotion during different interpretations of a Chopin piano prelude. Music Perception 19 (1), 87-120.

Sundberg, J. (1987). The science of the singing voice. Northern Illinois University Press.

Sundberg, J. (1988). Computer synthesis of music performance. In J.A. Sloboda (Ed.), Generative processes in music. The psychology of performance, improvisation, and composition (pp.52-69). Oxford: Clarendon Press.

Sundberg, J., Friberg, A., & Frydén, L. (1989). Common secrets of musicians and listeners: An analysis-by-synthesis study of musical performance. In P. Howell, R. West, & I. Cross (Eds.) Representing musical structure (pp. 161-197). London: Academic Press.

Timmers, R., Marolt, M., Camurri, A., & Volpe, G. (2006). Listeners' emotional engagement with performances of a Skriabin etude: An explorative case study. *Psychology of Music, 34 (4),* 481-510.

Timmers, R. (2007). Perception of music performance on historical and modern commercial recordings. Manuscript submitted for publication.

Todd, N. P. (1985). A model of expressive timing in tonal music. *Music Perception, 3,* 33-58.

Todd, N.P. (1992). The dynamics of dynamics: a model of musical expression. *Journal of the Acoustical Society of America, 91 (6),* 3540-3550.

Windsor, L.W., Desain, P., Penel, A., & Borkent, M. (2006). A structurally guided method for the decomposition of expression in music performance. *Journal of the Acoustical Society of America, 119 (2),* 1182-1193.

Appendix A: Emotional characterization of musical passages in JN

Text	Activity	Valence	Tension highpoint (Bar containing:)
Wie braust durch die Wipfel der heulende Sturm!	1	Neg	Sturm
Es klirren die Balken, es zittert das Haus!	1	Neg	Haus
Es rollet der Donner, es leuchtet der Blitz,	1	Neg	Blitz
Und finster die Nacht, wie das Grab!	0	Neg	Grab
Immerhin, immerhin,	0	Neg	immerhin
so tobt' es auch jüngst noch in mir!	1	Neg	Sturm
Es brauste das Leben, wie jetzo der Sturm,	1	Neg	Haus
Es bebten die Glieder, wie jetzo das Haus,	1	Neg	Blitz
Es flammte die Liebe, wie jetzo der Blitz,	1	Neg	Grab
Und finster die Brust, wie das Grab.	0	Pos	Sturm
Nun tobe, du wilder gewalt'ger Sturm,	1	Pos	Friede
Im Herzen ist Friede, im Herzen ist Ruh,	0	Pos	Glut
Des Bräutigams harret die liebende Braut,	1	Pos	ewigen
Gereinigt in prüfender Glut,	1	Pos	
Der ewigen Liebe getraut.	1	Neg	hole
Ich harre, mein Heiland! mit sehnendem Blick!	0	Neg	
Komm, himmlischer Bräutigam, hole die Braut,	0	Neg	
Erlöse die Seele von irdischer Haft.	0	Pos	Getön
Horch, friedlich ertönet das Glöcklein vom Turm!	0	Pos	ewigen
Es lockt mich das süße Getön	1	Pos	
Allmächtig zu ewigen Höhn.	1	Pos	
Es locht mich das süße Getön	1	Pos	
Allmächtig zu ewigen Höhn	1	Pos	
Alleluja! Alleluja!	0		

Appendix B: Emotional characterization of musical passages in DbdR

Text	Activity	Valence	Tension highpoint (Bar containing:)
Du bist die Ruh, der Friede mild,	0	Pos	Friede
Die Sehnsucht du und was sie stillt.	0	Pos	was
Ich weihe dir, voll Lust und Schmerz	1	Neg	Lust
zur Wohnung hier mein Aug und Herz, mein Aug und Herz.	1	Neg	Aug, Herz
Kehr ein bei mir, und schließe du	0	Pos	schließe
still hinter dir, die Pforten zu.	0	Pos	Pforten
Treib andern Schmerz, aus dieser Brust!	1	Neg	dieser
Voll sei dies Herz, von deiner Lust, von deiner Lust.	1	Neg	deiner, Lust
Dies Augenzelt, von deinem Glanz	2	Pos	
allein erhellt, O füll es ganz! O füll es ganz!	2 1	Pos Neg	erhellt, füll
Dies Augenzelt von deinem Glanz	2	Pos	
allein erhellt, O füll es ganz! O füll es ganz!	2 1 0	Pos Neg	erhellt, füll

Appendix C: Emotional characterization of musical passages in GaSp

Text	Activity	Valence	Tension highpoint (Bar containing:)
Meine Ruh ist hin, mein Herz ist schwer	0	Neg	Herz
(ich finde) ich finde sie nimmer, und nimmer mehr	1	Neg	nimmer
Wo ich ihn nicht hab, ist mir das Grab	0	Neg	mir
Die ganze Welt, ist mir vergällt.	0	Neg	ganze
Mein armer Kopf, ist mir verrückt,	1	Neg	
Mein armer Sinn, ist mir zerstückt	1	Neg	mir
Meine Ruh' ist hin, mein Herz ist schwer,	0	Neg	Herz
Ich finde sie nimmer, und nimmer mehr.	1	Neg	nimmer
Nach ihm nur schau ich, zum Fenster hinaus,	0	Neg	Fenster
Nach ihm nur geh ich, aus dem Haus.	0	Neg	Aus
Sein hoher Gang, sein' edle Gestalt,	0	Pos	
Seine Mundes Lächeln, seiner Augen Gewalt,	1	Pos	Augen
Und seiner Rede, Zauberfluß,	1	Pos	Zauberfluss
Sein Händedruck, und ach, sein Kuß!	2	Pos	Kuss
Meine Ruh' ist hin, mein Herz ist schwer,	0	Neg	Herz
Ich finde sie nimmer, und nimmermehr.	1	Neg	nimmer
Mein Busen drängt sich, nach ihm hin.	0	Pos	
Ach dürft ich fassen, und halten ihn	1	Pos	
Und küssen ihn, so wie ich wollt,	2	Pos	küssen
An seinen Küssen, vergehen sollt	2	Pos	
O Könnt ich ihn küssen, so wie ich wollt,	2	Pos	
An seinen Küssen, vergehen sollt!	2	Neg	Vergehen
O Könnt ich ihn küssen, so wie ich wollt,	2	Neg	
An seinen Küssen, vergehen sollt!	2	Neg	Vergehen
Meine Ruh' ist hin, mein Herz ist schwer,	1	Neg	Herz

Appendix D: Information about the recordings

Song	Performers	Ref	Source
JN	Susan Strong Orchestra	SS 07	'Schubert Lieder on Record I, 1898-1939', EMI Classics 5 66150 2, 1997
JN	Meta Seinemeyer Orchestra, Frieder Weismann	MS 28	Parlophon P 9663, matrix number 2-20726-1
JN	Susan Metcalfe-Casals Gerald Moore	SMC 37	'Schubert Lieder on Record II, 1929-1952', EMI Classics 5 66154 2, 1997
JN	Lotte Lehmann Paul Ulanowsky	LL 41	'Lotte Lehmann: Schubert', LYS 231-234, 1997
JN	Kathleen Ferrier Phyllis Spurr	KF 47	'The World of Kathleen Ferrier', Decca 430 096-2, 1990
JN	Elisabeth Schwarzkopf Edwin Fischer	ES 52	'Schubert: 12 Lieder, 6 Moments musicaux', EMI Classics 5 67494 2, 2000
JN	Elly Ameling Dalton Baldwin	EA 75	'Schubert Liederen', Philips 464 334-2, 1999
JN	Gundula Janowitz Irwin Gage	GJ 77	'Schubert Lieder', Deutsche Grammophon 453 082-2, n.d.
DbdR	Lilli Lehmann Fritz Lindemann	LiL 07	'Lilli Lehmann (1848-1929)', Lebendige Vergangenheit MONO 89185, 1999
DbdR	Elena Gerhardt Arthur Nikisch	EG 11	'Schubert Lieder on Record I, 1898-1939', EMI Classics 5 66150 2, 1997
DbdR	Lotte Lehmann Orchestra Odéon	LL 27	'Lotte Lehmann: Schubert', LYS 231-234, 1997
DbdR	Elisabeth Schumann Karl Alwin	ESu 32	'Elisabeth Schumann: Schubert Songs I', HMV COLH. 130, 1962
DbdR	Flora Nielsen Gerald Moore	FN 40s	HMV C 4057, matrix number 2EA 14375-1

Appendix 4 (continued)

Song	Performers	Ref	Source
DbdR	Dietrich Fischer-Dieskau Gerald Moore	DFD 69	'Franz Schubert Lieder II', Deutsche Grammophon 437 225-2, 1985
DbdR	Elly Ameling Dalton Baldwin	EA 76	'Schubert Liederen', Philips 464 334-2, 1999
DbdR	Gundula Janowitz Irwin Gage	GJ 76	'Schubert Lieder', Deutsche Grammophon 453 082-2, n.d.
GaSp	Emma Eames Henri Gilles	EE 11	'The Record of Singing I', EMI RLS 724, 1977
GaSp	Meta Seinemeyer Orchestra Frieder Weismann	MS 28	Parlophon P 9663 matrix number 2-20717-2
GaSp	Elisabeth Schumann Gerald Moore	ESu 36	'Elisabeth Schumann: Schubert Songs I', HMV COLH. 130, 1962
GaSp	Lotte Lehmann Ernö Balogh	LL 37	'Lotte Lehmann: Schubert', LYS 231-234, 1997
GaSp	Elisabeth Schwarzkopf Gerald Moore	ES 48	'Elisabeth Schwarzkopf: The unpublished EMI recordings, 1946-1952', Testament SBT 2172, 1999
GaSp	Elisabeth Schwarzkopf Edwin Fischer	ES 50	'Schubert: 12 Lieder, 6 Moments musicaux', EMI Classics 5 67494 2, 2000
GaSp	Elly Ameling Dalton Baldwin	EA 75	'Schubert Liederen', Philips 464 334-2, 1999
GaSp	Gundula Janowitz Irwin Gage	GJ 76	'Schubert Lieder', Deutsche Grammophon 453 082-2, n.d.