Acoustic Patterning of Speech
Its Linguistic and Physiological Bases

Phonetica 2000;57:95–112

Received: November 1, 1999
Accepted: February 21, 2000

Emotive Transforms

Johan Sundberg

KTH Voice Research Centre, Department of Speech Music Hearing,
KTH (Royal Institute of Technology), Stockholm, Sweden

Abstract

Emotional expressivity in singing is examined by comparing neutral and expressive performances of a set of music excerpts as performed by a professional baritone singer. Both the neutral and the expressive versions showed considerable deviations from the nominal description represented by the score. Much of these differences can be accounted for in terms of the application of two basic principles, grouping, i.e. marking of the hierarchical structure, and differentiation, i.e. enhancing the differences between tone categories. The expressive versions differed from the neutral versions with respect to a number of acoustic characteristics. In the expressive versions, the structure and the tone category differences were marked more clearly. Furthermore, the singer emphasized semantically important words in the lyrics in the expressive versions. Comparing the means used by the singer for the purpose of emphasis with those used by a professional actor and voice coach revealed striking similarities.

Introduction

Almost 100% of the world's population actively seek the occasion to listen to music. Also, music is often called the 'language of emotions', a poetic rather than precise description, that would simply allude to the fact that music often mediates strong emotional experiences to the listener. An important question emerges: how can music be understandable to almost anyone and how can it be so closely linked to emotions?

Obviously, the composer is responsible for a good deal of the emotional impact on the listeners. However, listening even to masterpieces of compositional art can be either wonderful or boring, depending on the performance. Thus, important contributions to musical experience derive also from the performer. Analysis of music performance should be a worthwhile object in the study of musical expressivity.

In our music performance research we have applied the analysis-by-synthesis strategy (fig. 1) [Sundberg, 1993]. The input is a music file containing the information given by the score, complemented with chord symbols and phrase and subphrase mark-

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/00/0574–0095
$17.50/0
Accessible online at:
www.karger.com/journals/pho

Johan Sundberg
KTH Voice Research Centre
Department of Speech Music Hearing, KTH
SE–10044 Stockholm (Sweden)
Tel. +46 8 790 7873, Fax +46 8 790 7854

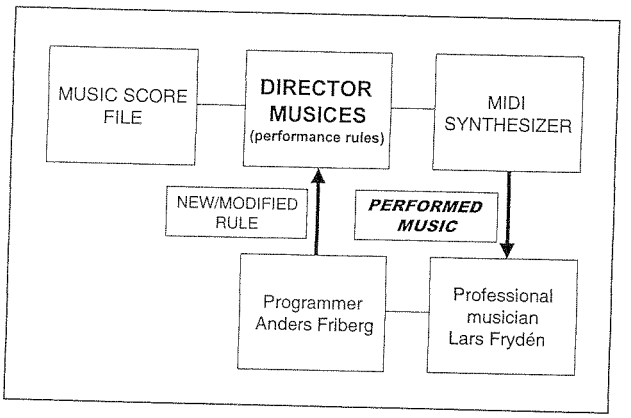


Fig. 1. The analysis-by-synthesis research paradigm. A music file is read by the Director Musices program, which identifies certain musical context and accordingly introduces departures from the nominal description of the piece provided by the score with respect of various available parameters, such as tone and pause duration, amplitude changes, vibrato parameters, and timbre.

ers. A performance program, Director Musices [Friberg, 1991, 1995], reads the music file. It contains a number of performance rules, which, depending on the musical context, insert micropauses and variations of amplitude, tempo, and vibrato. The output is a sounding performance. During the development the system has been continuously evaluated by an expert musician, Lars Frydén and his suggestions have been implemented in the Director Musices. In a sense, then, the Director Musices is a generative grammar of music performance, reflecting essential aspects of Frydén's musical competence. Independent corroboration of the rules has been gathered from listening tests [Friberg, 1995].

The Director Musices rules can be divided into two main groups according to their apparent purpose in music communication, differentiation rules and grouping rules. The differentiation rules enhance the difference between tone categories, such as scale tones, intervals, and note values. The grouping rules mark the musical structure, e.g. by inserting micropauses at structural boundaries. All rules are triggered by the musical context and thus ultimately reflect nothing but the musical structure [Palmer, 1989]. Apparently this implies that Director Musices performances cannot contain any emotional information.

Yet, Bresin and Friberg [1998] recently demonstrated that with appropriately chosen tempo and loudness, Director Musices is indeed capable of producing performances that differ in emotional expressivity. This is achieved by varying the selection of rules applied and the magnitudes of the rule effects. For example, a performance may sound happy, if the micropauses are made long, the tempo is quick, and the amplitude varies within and between tones. This finding is highly relevant to the issue of emotional transforms. It indicates that emotional expressivity can be derived directly from the music score simply by enhancing structural aspects of the piece.

The Director Musices grammar has been developed for instrumental music. In vocal music the emotional coloring of the performance seems particularly salient, and singers tend to succeed quite well in communicating the intended emotional information to the listener [Kotlyar and Morosov, 1976; Scherer, 1995]. In forced choice tests, listeners succeed in identifying intended emotions in about 80% of the cases, on average. Similar results have been found for performance on musical instruments [Gabrielsson, 1995; Juslin, 1997].

Table 1. Examples analyzed

Composer	Song title	Text	Abbreviation	Character
Folk tune	*Vi gå över daggstänkta berg,* bars 1–8	Vi gå över daggstänkta berg…	*Folk tune*	agitated
F. Schubert	*Erlkönig,* bars 72–79	Mein Vater, mein Vater,	*Mein Vater*	agitated
R. Schumann	*Dichterliebe VII,* bars 12–18	Wie Du auch strahlst…	*Wie Du auch*	agitated
R. Schumann	*Liederkreis XII,* bars 18–26	Und der Mond…	*Und der Mond*	agitated
G. Verdi	*Falstaff,* Ford's monologue, bars 24–31	Laudata sempre sia…	*Ford's Monologue*	agitated
G. Mahler	*Lieder eines fahrenden Gesellen,* song No. 3, bars 5–11	Ich hab' ein glühend Messer…	*Ich hab' ein*	agitated
F. Mendelssohn	*Paulus,* Aria No. 18, bars 5–13	Gott sei mir gnädig…	*Mendelssohn*	peaceful
F. Schubert	*Du bist die Ruh,* bars 8–15	Du bist die Ruh…	*Du bist die Ruh*	peaceful
F. Schubert	*Wanderers Nachtlied,* bars 3–14	Über allen Gipfeln ist Ruh…	*Wanderers Nachtlied*	peaceful
F. Schubert	*Nähe des Geliebten,* bars 3–8	Ich denke dein…	*Nähe des Geliebten*	peaceful
R. Schumann	*Dichterliebe VI,* bars 31–42	Es schweben Blumen und Englein	*Es schweben*	peaceful
R. Strauss	*Zueignung,* bars 21–29	Und beschworst darin die Bösen…	*Zueignung*	peaceful

The purpose of the present, exploratory investigation was twofold: (1) to examine some examples of emotive transforms in singing, and (2) to compare examples of the acoustic code used for adding expressivity in singing with some examples used for marking emphasis in speech. Results from two experiments will be reported, one concerning singing and one concerning speech.

Experiment I

Material collected for a previous investigation was used [Sundberg et al., 1995]. Håkan Hagegård, internationally well-known professional baritone, agreed to perform 17 excerpts of differing characters from the classical repertoire in two ways, as in a concert and with as little expression as possible.

The emotional expressivity of these examples was evaluated in two listening tests. In the first, a panel of expert listeners evaluated the difference in expressiveness between the neutral and expressive versions. Twelve excerpts, in which the expressive versions were perceived as particularly expressive as compared with the neutral versions, were selected for further analysis. In the second test, 5 experts classified the emotional quality of the expressive versions as either secure, loving, sad, happy, scared, angry, or hateful. The results, shown in table 1, implied that a majority perceived six examples as hateful or happy, i.e. agitated, and the remaining six as loving and secure, i.e. peaceful.

The recording were analyzed with regard to overall sound level, duration of tones, long-term average spectrum (LTAS), and fundamental frequency (F_0). A detailed description of the experiment and the analysis can be found elsewhere [Sundberg et al., 1995].

Tone duration has been found to play a prominent role in music performances. The measurement of tone duration was based on the result of a previous experiment with synthesized sung performances [Sundberg, 1989]. As its definition in sung performances is crucial to the results of this study, this experiment will be briefly reviewed. In speech research syllables are generally measured as in orthog-

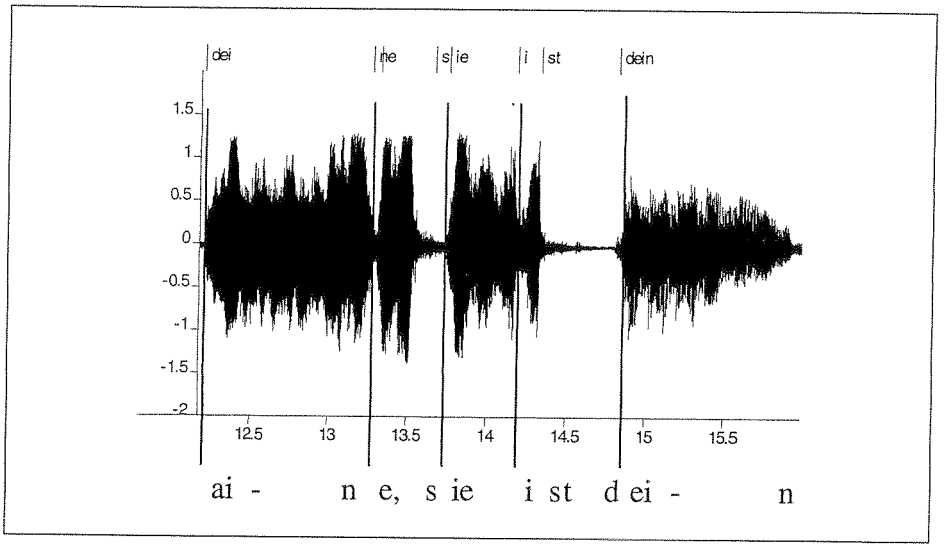


Fig. 2. Illustration of the segmentation of musical syllables according to the vowel onset-to-vowel onset criterion.

raphy, such that a CV syllable starts at the onset of the C. As a C is lengthened after a short V and shortened after a long V in singing, the duration of a C depends on the context. Hence, the duration of a CV syllable will differ depending on whether its duration is measured as in orthography or from vowel onset to vowel onset. In the synthesis experiment mentioned, the tones in a rhythmical example were assigned [la] and [la:] syllables segmented according to both these alternatives. The results revealed that a correct realization of the rhythmical structure was obtained only when syllables were segmented from vowel onset to vowel onset, as illustrated in figure 2. This demonstrates that syllable duration in singing should be measured from vowel onset to vowel onset.

Results

Figure 3 shows the mean overall sound level. On average, the agitated examples were louder than the peaceful examples, particularly in the expressive versions. The peaceful examples were softer in the expressive than in the neutral versions. Thus, the singer tended to enhance the loudness difference between the two example categories in the expressive versions. Figure 4 shows the short-term variability of sound level in terms of the mean of the time derivative, measured after a 20-Hz LP filter smoothing. The agitated examples showed a higher variability than the peaceful examples, particularly for the expressive versions.

Figure 5 shows the tempo, measured as the mean number of shortest note values per second. The agitated examples were sung at a faster tempo than the peaceful examples, which, in turn, were sung slower in the expressive than in the neutral version. Also in this case the singer increased the difference between the two example categories in the expressive versions.

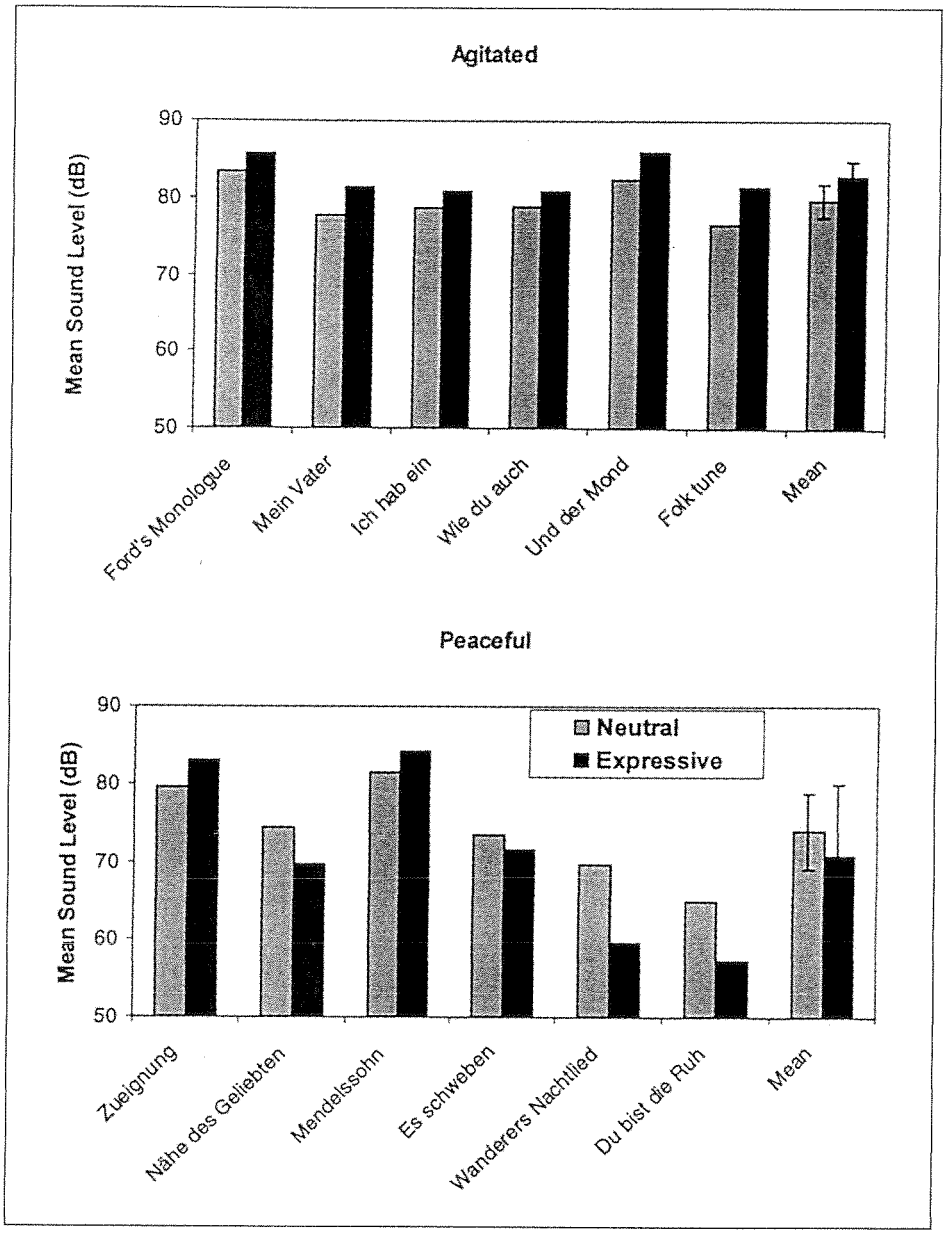


Fig. 3. Mean overall sound level, L_{eq} in the exerpts examined. Gray and white columns refer to neutral and expressive versions. Results for examples with an agitated and a peaceful emotional character. The bars in the two rightmost columns represent ±1 standard deviation.

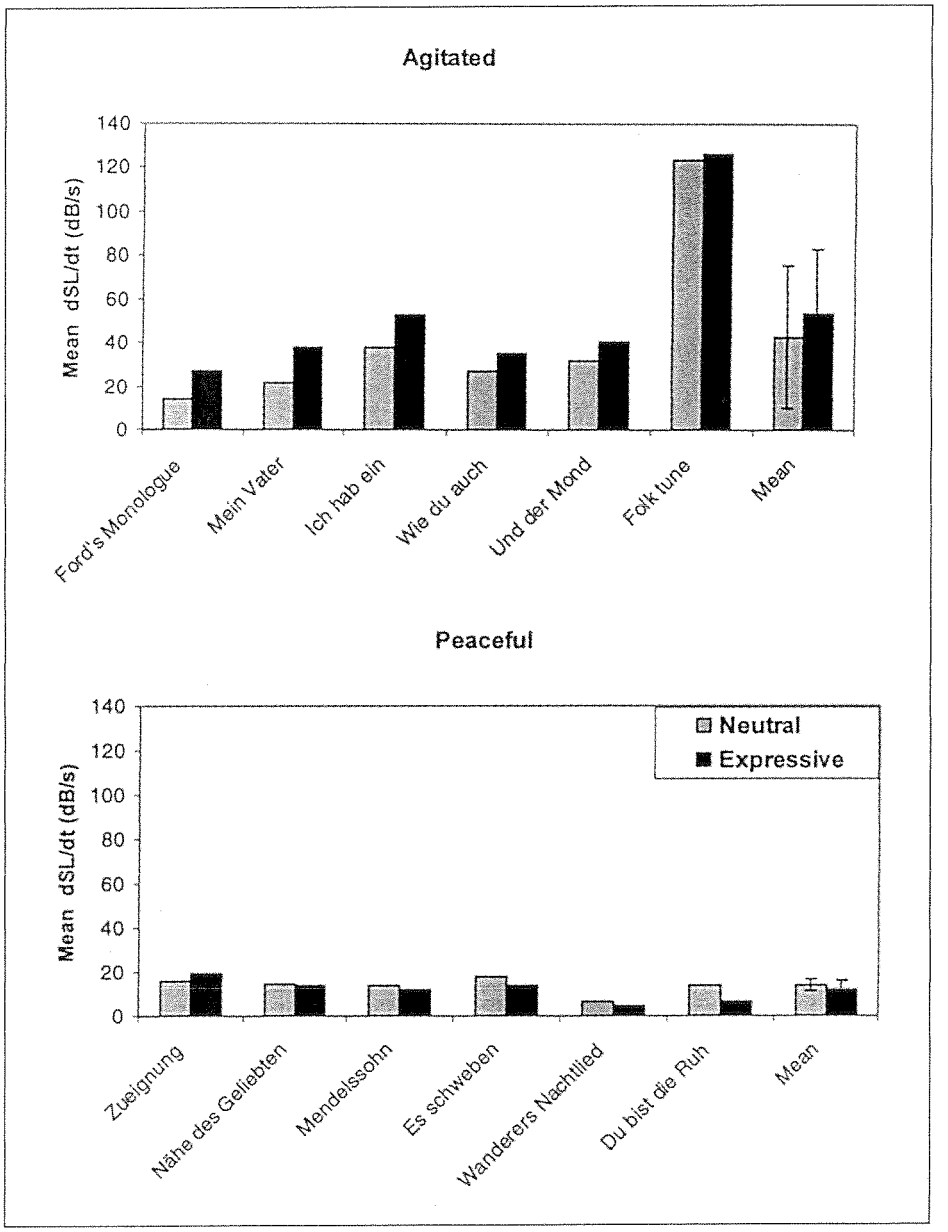


Fig. 4. Short-term variability of loudness. The columns represent the mean of the time derivative of the overall sound level, measured after a 20-Hz LP filter smoothing. Gray and white columns refer to neutral and expressive versions. Results for examples with an agitated and a peaceful emotional character. The bars in the two rightmost columns represent ±1 standard deviation.

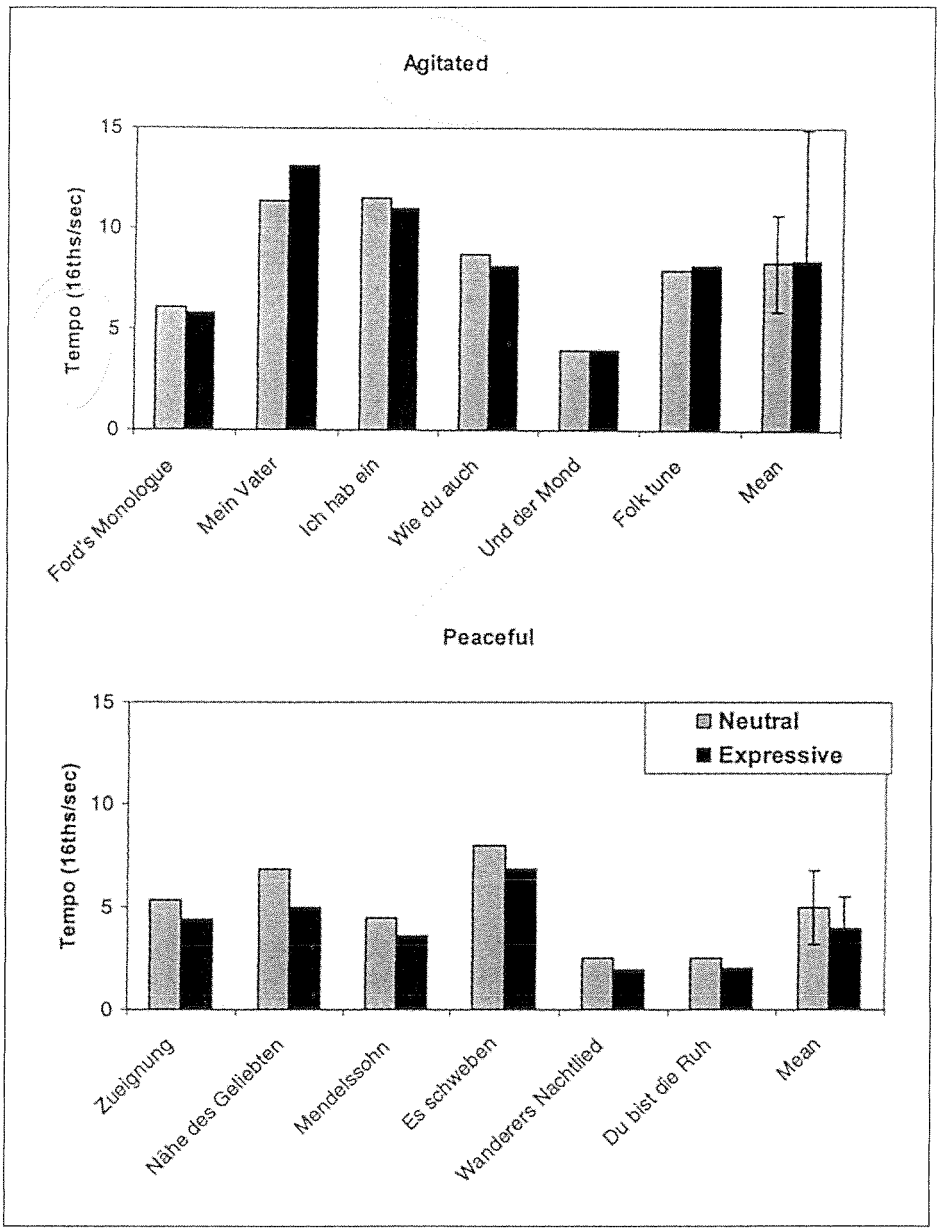
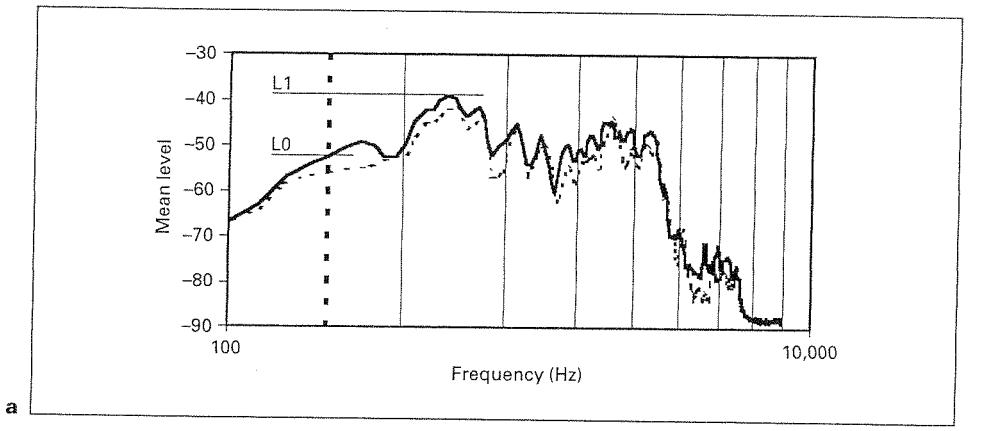


Fig. 5. Mean tempo, measured as the mean number of shortest note values per second. Gray and white columns refer to neutral and expressive versions. Results for examples with an agitated and a peaceful emotional character. The bars in the two rightmost columns represent ±1 standard deviation.



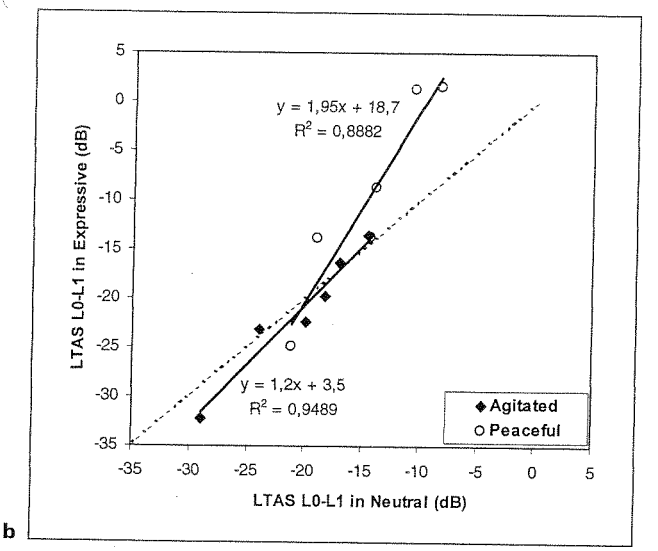


Fig. 6. a LTAS of the neutral and expressive performances of an example (dashed and solid curves). The dashed line represents the mean F_0. The LTAS levels at the mean of F_0 (L0) and at the highest peak (L1) are shown for the expressive version. **b** Dominance of the fundamental, estimated as the level difference LTAS L0–L1, for the neutral and expressive versions of the agitated and peaceful examples (filled and open symbols).

The level difference between the fundamental and the first formant changes with glottal adduction and thus reflects an aspect of phonation [Sundberg, 1987]. This aspect was studied from LTAS of each excerpt. The highest peak in an LTAS of vocal material roughly corresponds to the mean level at the mean of F_1 (fig. 6a). From such spectra the level was determined at the frequency corresponding to the average F_0 of the example. This level, henceforth LTAS L0, and the level of the main peak of the LTAS, L1, are marked for the expressive version in the example shown in the figure. As illustrated in figure 6b the difference LTAS L0–L1 tended to be greater in the peaceful examples, particularly in the expressive versions. In other words, the fundamental was mostly more dominant in the expressive versions of the peaceful examples. Thus, the singer apparently performed the peaceful examples with less glottal adduction than the agitated examples.

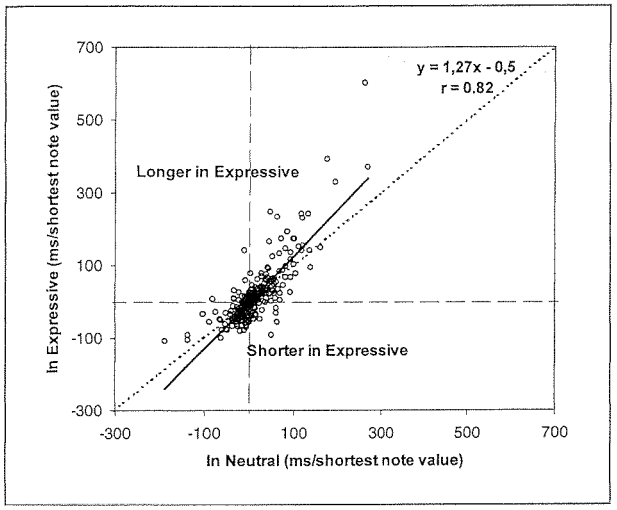


Fig. 7. Deviations from nominal duration in the expressive and neutral versions for all tones in all excerpts. The dotted line represents the case that the deviations were identical in both versions. The solid line represents the trendline, equation and correlation shown in the upper right corner.

According to Fónagy [1976], articulatory movements play a prominent role in the emotional coloring of speech. Formant frequency transition time was analyzed for some agitated and peaceful examples. Because of the comparatively high F_0, measurement was difficult in many cases. In the cases where reliable data were available, surprisingly small differences were found. However, a more detailed analysis is required before any conclusions can be drawn.

The score specifies durational relations between the various tones in nominal terms rather than as they are realized in a performance. Hence, a comparison between nominal and performed durations is interesting. Figure 7 compares normalized deviations from nominal durations in the expressive and neutral versions of all excerpts in terms of the mean lengthening per shortest note value. Had the tones in the neutral versions not deviated from nominal duration, all data points would have clustered around the vertical axis in the diagram. Instead, they are scattered between −200 and +300 ms. This indicates that the tone durations deviated considerably from nominal in the neutral versions. The solid line represents the best linear fit. Had the tones in the expressive versions departed from nominal durations as much as in the neutral versions, the trend line would have fallen upon the diagonal. Instead it shows a slope of 1.27. This indicates that the singer made similar departures from nominal duration in the expressive versions as compared to the neutral versions, but the departures were greater in the expressive versions. Thus, some of the deviations from nominal durations, that the singer used in the expressive versions, transpired also to his neutral versions; similar observations have been made in performance of instrumental music [Palmer, 1989].

Music structure is hierarchical; small groups of tones, musical gestures, join to form greater groups of tones, subphrases, which join to make still greater groups of tones, phrases, etc. This hierarchy is reflected in performance; typically, musical gestures are terminated with a micropause, and subphrases and phrases are initiated by an accelerando from a slow tempo and terminated with a rallentando (a slowing down of the tempo). These characteristics have been implemented in a model illustrated in

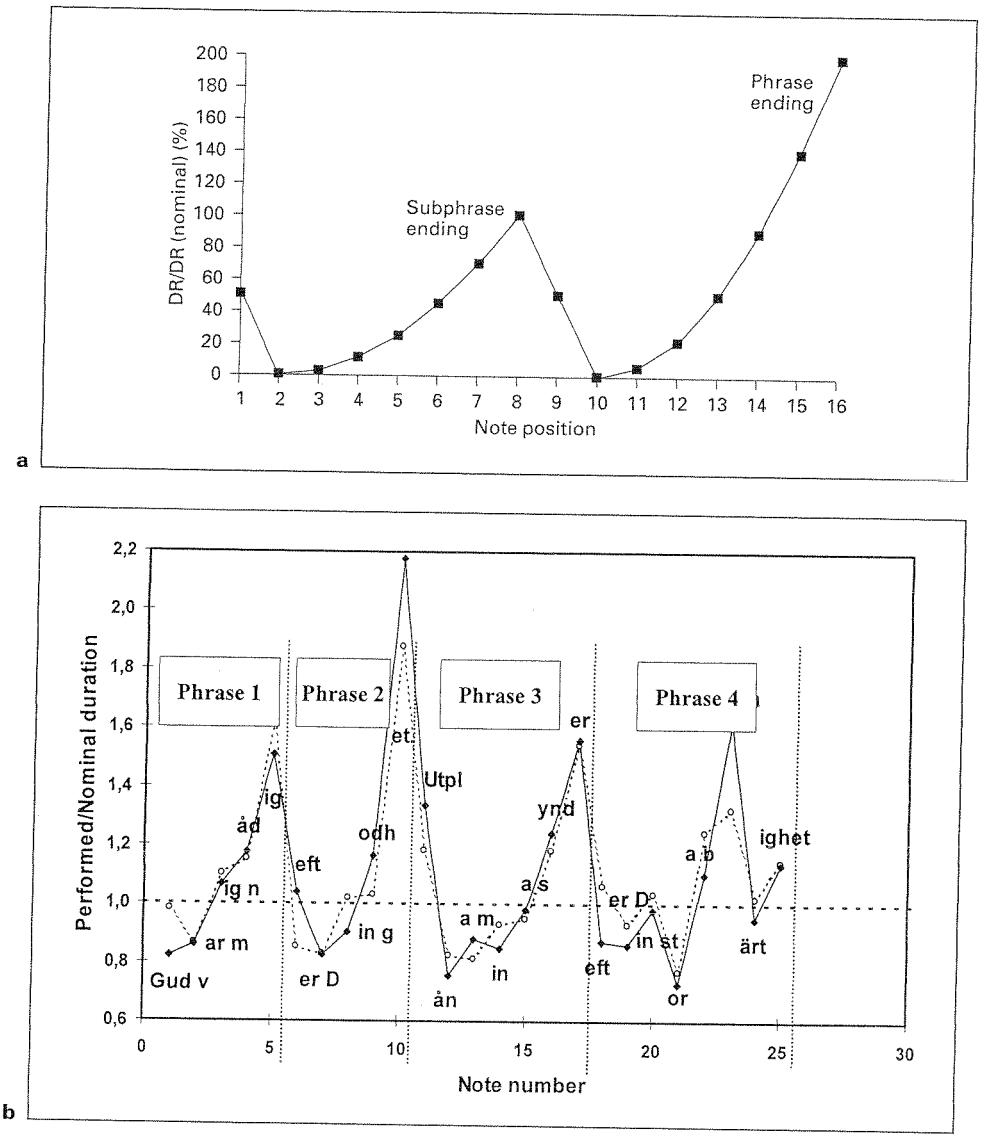


Fig. 8. **a** Model for the tempo curve, expressed as the ratio between performed and nominal duration (DR), for subphrases and phrases implemented in the Director Musices program [from Friberg and Sundberg, 1999]. **b** Performed-to-nominal duration ratios in the singer's neutral and expressive versions of the *Mendelssohn* example. The nominal durations were calculated on the basis of the mean duration of the shortest note value in the entire example.

figure 8a [Friberg and Sundberg, 1999]. Figure 8b shows a typical sung example in terms of deviations from the nominal durations, which were calculated on the basis of the mean duration of the shortest note value in the entire example. The example illustrates the observation above that the deviations were similar but often slightly greater in the expressive than in the neutral versions.

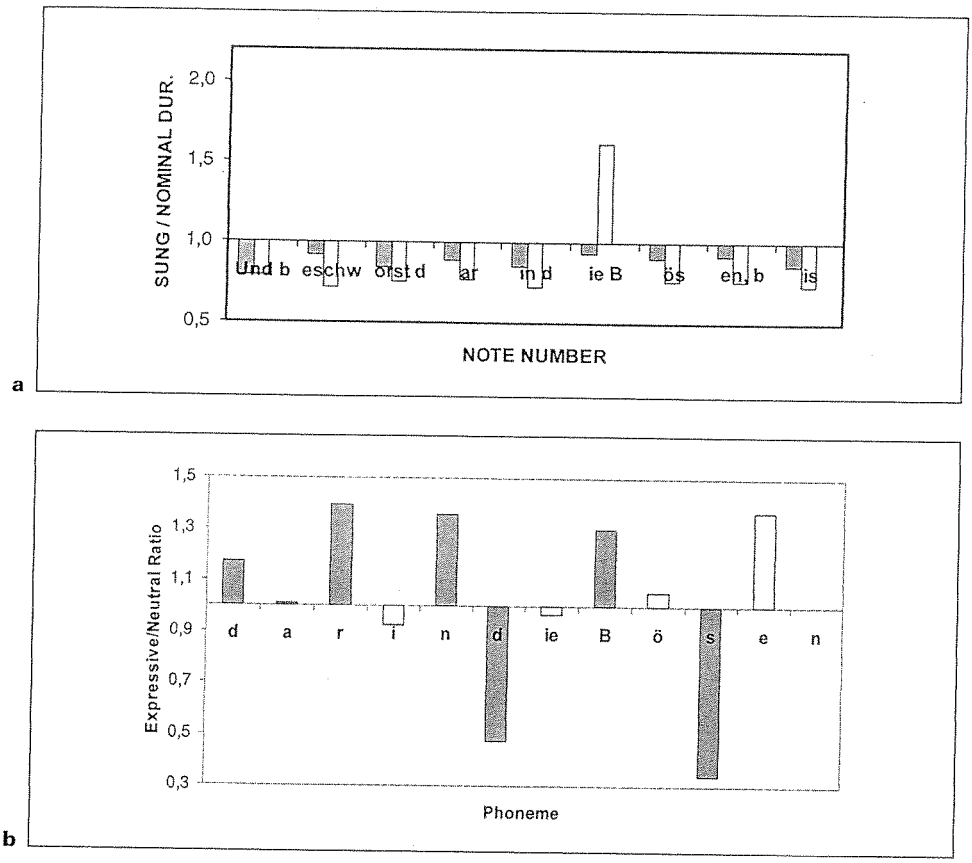
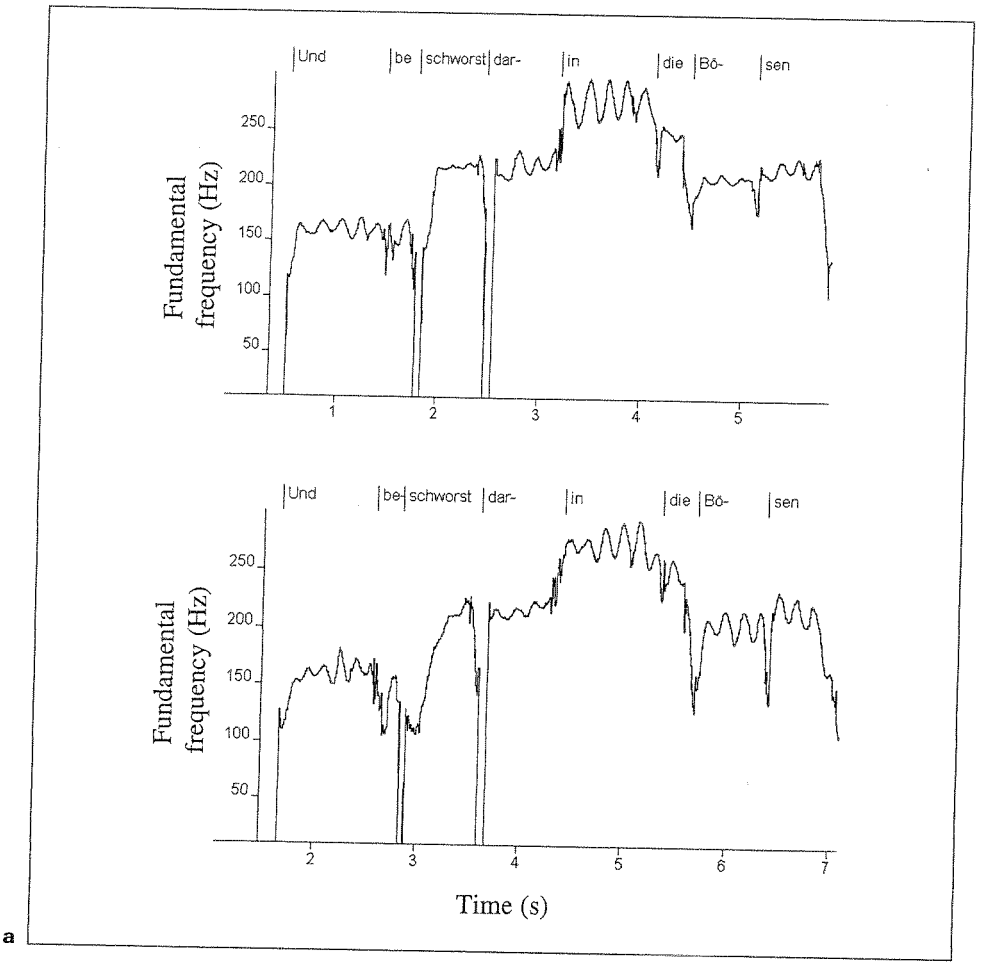


Fig. 9. a Performed-to-nominal duration ratios for the syllables of the *Zueignung* example. The gray and white columns refer to the neutral and expressive versions. **b** Phoneme duration ratio between the expressive and neutral versions for a section of the *Zueignung* example. Vowels are represented by white columns.

Vocal performers often seem to emphasize words that they perceive as particularly important for semantic reasons. Emphasized words in the material were identified by an informal listening test. The singer seemed to use different methods for marking emphasis.

One method was to lengthen the stressed syllable of the emphasized word. The performed/nominal duration ratios were determined for all syllables in the neutral and expressive versions of all excerpts. These ratios were then compared between the neutral and expressive versions for each syllable. In 34 cases the performed/nominal duration ratio for a syllable was more than 20% greater in the expressive than in the neutral version. Of these, 16 lengthenings occurred for syllables that appeared in a stressed position in the bar. Thus, in these cases the singer lengthened the stressed syllable in emphasized words.

Another method to emphasize words was observed almost as frequently. In 18 of the 34 cases just mentioned the lengthenings occurred on syllables that appeared on the

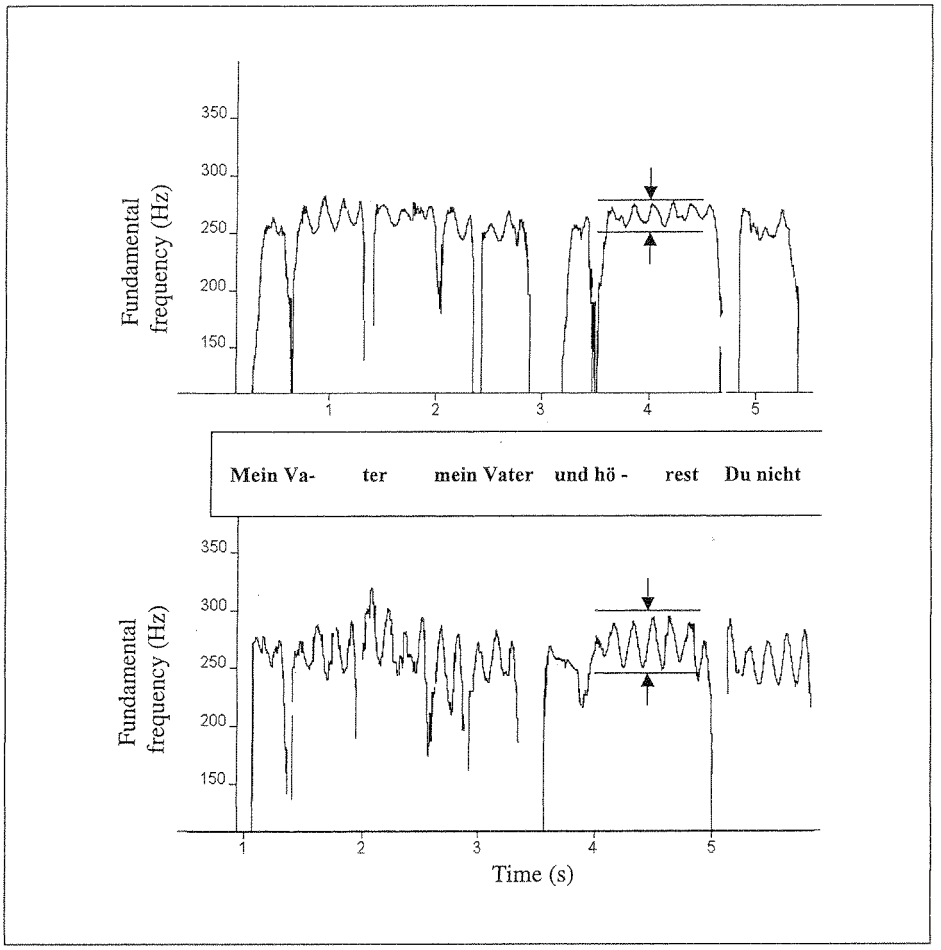


a

Fig. 10. a F_0 curves for the neutral and expressive versions of the *Zueignung* example. **b** F_0 curves for the neutral and expressive versions of the *Mein Vater* example.

upbeat of the emphasized syllable of the word. Thus, in these cases the lengthening occurred in an unstressed position in the bar, i.e. on the syllable preceding the stressed syllable of the emphasized word. As a result, the stressed syllable of the emphasized word was somewhat delayed. This phenomenon might be called the *emphasis by delayed arrival.*

Figure 9a shows a typical example. Here, the word 'Bösen' (evil) was perceived as emphasized, which seems logical from a semantic point of view. Although appearing in an unstressed upbeat position in the bar, the syllable (d)'ie B'(ösen) was clearly lengthened, while the syllable (B)'ös'(en) was slightly shortened in the expressive version. Figure 9b shows the phoneme duration ratio between expressive and neutral for this section of the text. It can be seen that the lengthening concerned the consonant [b] rather than the vowel preceding it. Several similar examples were observed in the material.



10b

Other events that seemed typically associated with perceived emphasis consisted of specific pitch patterns. Figure 10a shows an example. In the neutral version of this peaceful excerpt, F_0 changed quickly between the tones while in the expressive version, long and wide ascending pitch glides occurred on 'Und' and in '(b)eschw(orst)'. Such pitch glides did not characterize all expressive versions. In the agitated example shown in figure 10b, pitch glides can be seen in the neutral rather than in the expressive version on the phrase-initial words 'Mein' and 'und'. In the expressive version the pitch curve changed more abruptly. In addition, the figure illustrates that the extent of the vibrato modulation in agitated examples was much greater in the expressive version. This is in agreement with the observation that the extent and rate of the vibrato are important in signaling the emotion of fear in music performances [Gabrielsson and Juslin, 1996].

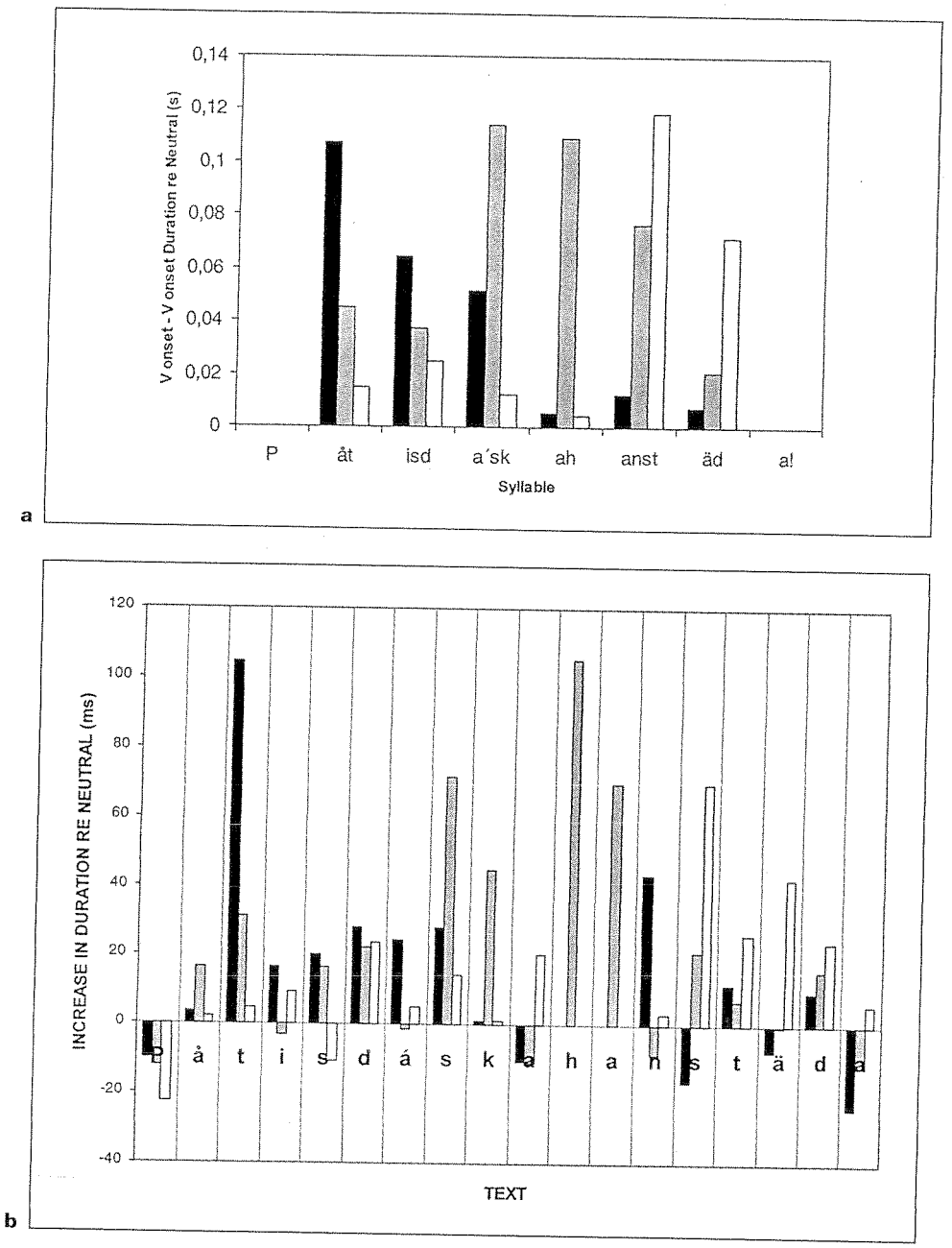


Fig. 11. a Syllable duration differences between the neutral and emphasized cases for one of the spoken sentences. Black, gray and white columns refer to the versions where the words 'tisdag', 'ska' and 'städa', respectively, were emphasized. **b** Phoneme duration differences between the neutral and emphasized cases for one of the sentences. Black, gray and white columns refer to the versions where the words 'tisdag', 'ska' and 'städa', respectively, were emphasized.

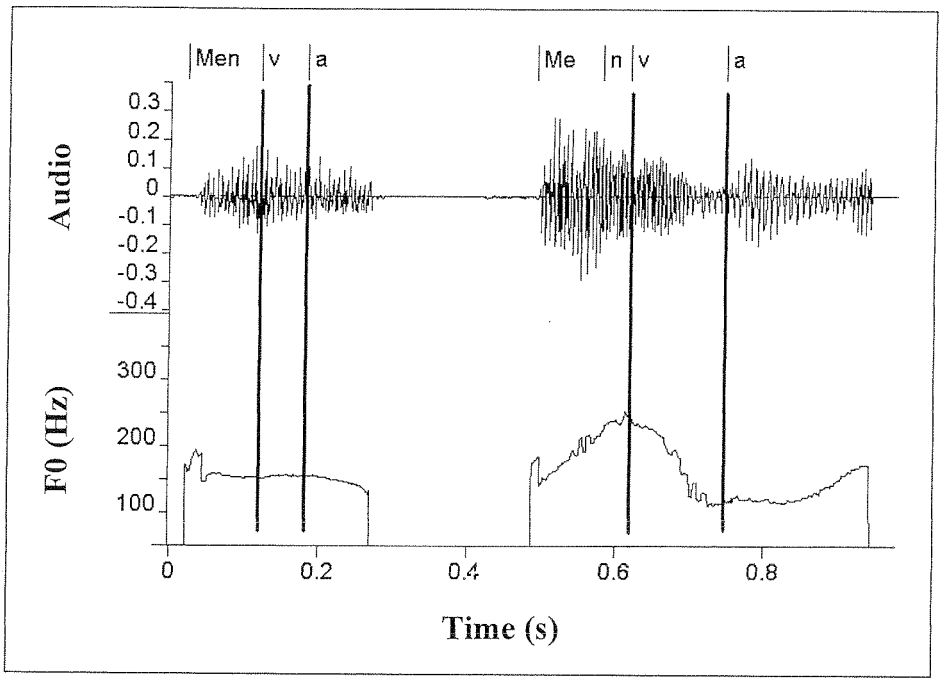


Fig. 12. Audio and F_0 curves for the spoken words 'men vad' pronounced in a neutral way (left) and with emphasis on the word 'vad' (right).

Experiment II

One method that the singer used to emphasize words was to lengthen the stressed syllable of the emphasized word. Fant et al. [1999] recently found a strong correlation between perceived stress and vowel duration in speech. Thus, in these cases the singer used the same code for emphasis that can be used in speech.

Several examples were also found of emphasis by delayed arrival, i.e. lengthening the unstressed tone preceding the stressed syllable of an emphasized word. To find out if this principle is also applied in speech, it is necessary to measure the duration of syllables defined in the same way as in singing, i.e. from vowel onset to vowel onset. This was realized by a simple experiment.

A female actor and hightly experienced voice coach was asked to read a set of short Swedish sentences, first six times in a neutral way, and then three times emphasizing one of the different words in the sentence. Vowel onset to vowel onset syllable duration was measured, and means were calculated for each condition. The syllable durations observed for the neutral version were used as a reference, such that lengthenings and shortenings were calculated relative to the neutral version.

Results

Figure 11a shows the differences between the neutral and emphasized cases for one of the sentences. Not only the emphasized syllables, but also the syllable preceding them were clearly lengthened. For instance, when 'tisdag' (Tuesday) was emphasized, not only 'isd' but also the preceding syllable 'å t' showed increased duration. A seg-

ment duration analysis of the same sentence showed that consonants preceding emphasized vowels tended to be lengthened (fig. 11b). For example, the consonant [t] was clearly lengthened in the version emphasizing the word 'tisdag'. Similarly, when the word 'städa' (tidy) was emphasized, the initial consonants [s] and [t] were lengthened. Heldner [1996] has made similar observations. Thus, also in this case, we find a similarity of emphasis markers in speech and singing.

The same spoken material was analyzed also with respect to F_0 contours. Figure 12 compares a neutral version of the sentence 'Men vad (betyder det?)' (But what does it mean?) with the version where 'vad' (what) was emphasized: in the neutral version the duration of the consonant [v] was short and the pitch contour is part of an overall gently descending glide. In the emphasized version it was produced with a marked pitch gesture. This is similar to the pitch contour observed in the same consonant in 'Und beschworst'. This shows that at least part of the emphasis markers used for voiced consonants in singing can be found also in speech.

General Discussion and Conclusions

Above some examples of emotive transforms in singing have been discussed. Although synthesis experiments are certainly needed to test the generality of the observations made, the results still invite to some speculation.

This study relies heavily on the professional competence of the singer and the actor. Their expertise is to detect the emotional character of a text and to realize it in an understandable way to listeners. Interestingly, composers often leave most of the emotional interpretation to the performer. In many of the examples considered here, the composer's instruction is limited to hints for tempo, e.g. 'Rather slow'. Therefore, performer's ability to sense the emotional character of the text and the music as well as to realize it in an intelligible way is crucial. The striking similarities between the expressive means used in singing and in speech suggest that most listeners correctly perceive the emotional transforms. This must limit the variability of sung performances. For example, listeners are likely to react negatively, if a singer performs a peaceful song in an agitated way.

In this study, most observations have concerned examples of the differentiation principle, while the only example of the grouping principle mentioned was phrasing. The underlying assumption was that much of the essence of emotional transforms seemed likely to occur in the differentiation of tone categories. It seems that the marking of the hierarchical structure is in a sense a more basic aspect of music performance than the emotional coloring. For example, the singer marked the phrases almost as clearly in the neutral as in the expressive versions. Yet, a performer's emotional coloring of a song can be expected also to affect the marking of tone groups. Thus, the sound level changes reflecting the harmonic progressions were often greater in the expressive than in the neutral versions.

The code used by our singer subject to color his expressive versions emotionally is largely similar to that used in speech. The singer sang the agitated examples louder, with greater amplitude variation and at a faster tempo than the peaceful examples. These three acoustic characteristics have been found typical of the expression of activity during speech [Scherer, 1995]. Likewise, similarities were found with regard to how emphasis was signaled in singing and speech. For example, the singer lengthened

stressed syllables in emphasized words. As mentioned, this code for signaling emphasis has been found also in speech [Fant et al., 1999]. Furthermore, we found examples of emphasis by delayed arrival both in the singer's expressive versions and in the actor's speech.

The principle of delayed arrival implies lengthening of an unstressed syllable preceding a stressed syllable. The lengthening was found to concern not only the vowel but also the consonant of the unstressed syllable. If sung syllables were defined as in orthography, such lengthened consonants would belong to the stressed syllable, so the lengthening would occur on the stressed syllable. However, as mentioned before, this definition of syllables does not apply to singing. Yet, a particularly interesting case would be when the unstressed syllable preceding the stressed contained only a vowel. (John Kingston is gratefully acknowledged for pointing out the relevance of this case.) Only one such case occurred in the entire material, 'die Augen...' in the example *Es schweben.* Also in this case the syllable 'ie' was lengthened. Although more such examples should be analyzed, our results clearly suggest that delayed arrival is a useful emphasis marker in some sung contexts.

The level of the fundamental tended to be higher in the peaceful than in the agitated examples. This suggests that the singer varied the voice source depending on the emotional character of the song, using more glottal adduction in the agitated than in the peaceful examples. It is tempting to speculate about the reason for this. In speech a high degree of glottal adduction is typically used for loud phonation at high pitches, such as in shouting or screaming, while a low degree of adduction is common in soft voice. Hence, an increase of glottal adduction in agitated examples and a decrease in peaceful examples is likely to contribute to the emotional expressivity. A similar reasoning seems applicable to the difference in tempo, sound level, and sound level variability in the two types of examples.

What are the emotive transforms? As demonstrated by Bresin and Friberg [1998], essential contributions to emotional expressivity in piano performances seem to derive from the application of the two performance principles *grouping* and *differentiation.* This indicates that the marking of the hierarchical structure and the enhancing of the differences between tone categories contribute to emotional expressivity.

As might be expected, examples of the grouping and differentiation principles were also found in the sung performances analyzed here. Moreover, the effects were mostly slightly greater in the expressive than in the neutral performances. This observations has also been made for performances of instrumental music [Palmer, 1989]. Thus, a clearer differentiation of tone categories and a clearer marking of structure seem to be components in the emotive transforms.

The singer seemed to apply the differentiation principle also with respect to the emotional coloring of the performance. For example, he sang most of the agitated examples louder than the peaceful examples and in the expressive versions he increased this difference by singing the expressive versions louder than the neutral versions. Similarly, the sound level variability was on average greater in the agitated than in the peaceful examples and he enhanced this difference in the expressive versions. Thus a clearer marking of the acoustic code used for emotional expressivity seems to be a component of the emotive transforms.

The emphasizing of semantically important words seemed to be typical of the expressive versions of the sung examples. This can also be seen as a case of differentiation, although based upon semantics rather than musical structure. Thus, the princi-

ples of grouping and differentiation seem highly relevant to emotive expressivity. Both can be found in speech [Lindblom, 1979; Diehl, 1991] and also in other forms of communication. Indeed, they may be quite basic to human communication in general [Carlson et al., 1989].

Fónagy [1962, 1976] launched the idea of glottal and articulatory movement as a leading principle underlying expressivity in speech. We found examples of variation in glottal parameters, such as pitch and the dominance of the fundamental. Thus, our results support the assumption that expressive transforms are closely related to glottal factors.

Two questions were asked in the introduction: How can music be emotionally expressive, and why is music so widely appreciated? Our observations have shown that the principles of grouping and differentiation seem instrumental in producing emotive transforms in singing. They seem equally relevant to speech. In this sense, music is not special. Moreover, we have found many examples of identity or similarity in the code used in music and speech for the purposes of grouping, differentiation, and emotional coloring. This should make music understandable and possible to interpret in emotional terms to almost anyone who understands speech.

References

Bresin, R.; Friberg, A.: Emotional expression in music performance: synthesis and decoding. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 4, pp. 85–94 (1998).

Carlson, R.; Friberg, A.; Frydén, L.; Granström, B.; Sundberg, J.: Speech and music performance: parallels and contrasts. Contemp. Music Rev. *4:* 389–402 (1989).

Diehl, R.: The role of phonetics within the study of language. Phonetica *48:* 120–134 (1991).

Fant, G.; Kruckenberg, A.; Liljencrants, J.: Prominence correlates in Swedish prosody. Int. Congr. Phonet. Sci. 99, San Francisco 1999, vol. 3, pp. 1749–1752.

Fónagy, I.: Mimik auf glottaler Ebene. Phonetica *8:* 209–219 (1962).

Fónagy, I.: La mimique buccale. Phonetica *33:* 31–44 (1976).

Friberg, A.: Generative rules for music performance: a formal description of a rule system. Computer Music J. *15:* 56–71 (1991).

Friberg, A.: A quantitative rule system for musical performance; doct. diss. KTH, Stockholm (1995).

Friberg, A.; Sundberg, J.: Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. J. Acoust. Soc. Am. *105:* 1469–1484 (1999).

Gabrielsson, A.: Expressive intention and performance; in Steinberg, Music and the mind machine, pp. 35–47 (Springer, Berlin 1995).

Gabrielsson, A.; Juslin, P.: Emotional expression in music performance: between the performer's intention and the listener's experience. Psychol. Music *24:* 68–91 (1996).

Heldner, M.: Phonetic correlates of focus accents in Swedish. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 2, pp. 33–36 (1996).

Juslin, P.: Emotional communication in music performance: a functionalist perspective and some data. Music Percept. *14:* 383–418 (1997).

Kotlyar, G.M.; Morosov, V.P.: Acoustical correlates of the emotional content of vocalized speech. Sov. Physics-Acoustics *22:* 208–211 (1976).

Lindblom, B.: Final lengthening in speech and music; in Gårding, Bruce, Bannert, Nordic prosody. Travaux de l'Institut de Linguistique de Lund, No. 13, pp. 85–101 (1979).

Palmer, C.: Mapping musical thought to music performance. J. exp. Psychol. *15:* 331–346 (1989).

Scherer, K.: Expression of emotion in voice and music. J. Voice *9:* 235–248 (1995).

Sundberg, J.: Synthesis of singing by rule; in Mathews, Pierce, Current directions in computer music reasearch. System Development Foundation Benchmark Series (MIT Press, Cambridge 1989). With sound examples on CD ROM, 45–55 and 401–403.

Sundberg, J.: How can music be expressive? Speech Commun. *13:* 239–253 (1993).

Sundberg, J.; Iwarsson, J.; Hagegård, H.: A singer's expression of emotions in sung performance; in Fujimura, Hirano, Vocal fold physiology: voice quality and control, pp. 217–232 (Singular Publishing Group, San Diego 1995).

Sundberg, J.: The Science of the Singing Voice (Northern Illinois University Press, De Kalb, Illinois 1987).