Replicability and accuracy of pitch patterns in professional singers

Sundberg, J. and Prame, E. and Iwarsson, J.

journal: STL-QPSR
volume: 36
number: 2-3
year: 1995
pages: 051-062

http://www.speech.kth.se/qpsr
Replciability and accuracy of pitch patterns in professional singers

Johan Sundberg, Eric Prame and Jenny Iwarsson

Abstract
Pitch is a factor of prime relevance in music. It is typically perceived in terms of the scale tones categories, but most music listeners are also able to decide whether or not the pitch is in tune. In some instruments, such as the singing voice, pitch can be continuously varied. In this investigation, professional singers' voice fundamental frequency was measured in 10 singers' performances of a song recorded on commercially available phonograms. Fundamental frequency was determined as the mean value between the turning points of the vibrato cycles. A panel of expert listeners judged which tones were out-of-tune. The singers' reproducibility of fundamental frequency values was examined by comparing the tones which had identical text in two verses of the song; the mean difference over these tones as sung by all singers was found to be 7 cent. The mean F0 of tones, which no jury member found out-of-tune, agreed within about ±7 cents for most tones and departed significantly from equally tempered tuning for some tones. It is concluded that singers generally have to match the target F0 within ±7 cent and that "in-tune" is not equal to "in accordance with the equally tempered tuning".

Introduction
The human voice offers fewer constraints to its users than any other musical instrument. All other instruments contain at least fixed resonances and several also offer the player only a restricted choice of fundamental frequencies and an ensemble of preset degrees of loudness. In the human voice, by contrast, resonances, oscillator characteristics, and loudness can all be continuously varied. This raises exceptionally high demands on the control of these parameters on the part of the singer. Therefore, singing represents an interesting object of study regarding how accurately man can control a complex sound generating system.

One of the aspects of major relevance to the esthetic quality of singing is fundamental frequency control. As music listeners we seem to have very precise demands as to the singer's matching target pitches, i.e., if they stay "in-tune" or not. This investigation examines professional singers' fundamental frequencies. Three basic questions will be considered: (1) How accurately do professional singers replicate their fundamental frequencies? (2) How accurately do they realise target fundamental frequencies? and (3) Which are these target fundamental frequencies?

Method
The material analysed consisted of commercially available CD recordings of ten internationally renowned singers performing Franz Schubert's Ave Maria. Twenty-five long tones of each of these recordings were analysed as part of co-author EP's project on acoustic vibrato characteristics. Henceforth, these tones will be referred to in terms...
of the numbers shown in Figure 1. Recording data are given in Table 1. Most recordings had piano or harp accompaniment. The text was sung in German and Latin in 7 and 3 recordings, respectively.

![Ave Maria musical notation]

**Figure 1.** Franz Schubert's song *Ave Maria* with the tone numbers used in the present investigation.

Using a Kay Elemetrics Sonagraph 5500, the fundamental frequency of each of the 25 long tones were determined from a strong high overtone. For each tone the frequency at the turning points of each vibrato cycle were measured throughout the tone. Then, a running linear mean F0 was calculated of each subsequent pair of turning point values. This average corresponds to the pitch perceived (Sundberg, 1978: Shonle & Horan, 1980). Some of the tones in some recordings could not be measured because of too low recording level. A total of more than 5000 turning points pertaining to 237 tones were measured. Then, the mean F0 over all measured vibrato cycles of each tone were determined.

The tuning of the accompaniment varied greatly between the recordings, corresponding to an F0 variation of the pitch of A4 from 439.5 Hz to 448 Hz, or 32 cent. Thus, it was necessary to relate the singers’ F0 data to the tuning of the accompaniment. This tuning was determined by a professional musician, a violinist. The first bar of the accompaniment was repeatedly played and the musician adjusted
the frequency of a variable tone with triangular pulse shape such that it agreed with the pitch of the root of the accompaniment chord. He made three attempts for each of the ten recordings. The results varied within an interval no greater than 0.5 Hz, or 2 cent. These three frequency values were averaged for each recording. The frequencies of the scale tones were computed according to the equally tempered tuning, using this frequency as the reference for the singer’s F0 values. Henceforth, the singers’ intonations will be compared with these scale tone frequencies, referred to as the equally tempered tuning (ETT).

Table 1. Data on the ten analysed recordings of Franz Schubert’s song Ave Maria. Tempo is given in terms of the mean duration in $s$ of a quaver note.

<table>
<thead>
<tr>
<th>Nr</th>
<th>Class.</th>
<th>Key</th>
<th>Age</th>
<th>Year</th>
<th>Langu.</th>
<th>Accomp.</th>
<th>Tempo</th>
<th>Verses</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sopr</td>
<td>Bb4</td>
<td>37</td>
<td>1975</td>
<td>Germ</td>
<td>Piano</td>
<td>1.95</td>
<td>3</td>
<td>Verse 3 softer than verse 1 and 2</td>
</tr>
<tr>
<td>2</td>
<td>Sopr</td>
<td>Bb4</td>
<td>42</td>
<td>1990</td>
<td>Germ</td>
<td>Orchestra</td>
<td>2.18</td>
<td>3</td>
<td>Different arrangement of verses 1 and 2</td>
</tr>
<tr>
<td>3</td>
<td>Sopr</td>
<td>Bb4</td>
<td>41</td>
<td>1984</td>
<td>Latin</td>
<td>Harp</td>
<td>1.80</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Sopr</td>
<td>Bb4</td>
<td>41</td>
<td>1988</td>
<td>Germ</td>
<td>Piano</td>
<td>2.13</td>
<td>3</td>
<td>Verse 3 softer than verse 1 and 2</td>
</tr>
<tr>
<td>5</td>
<td>Sopr</td>
<td>Bb4</td>
<td>26</td>
<td>1971</td>
<td>Germ</td>
<td>Piano</td>
<td>2.10</td>
<td>3</td>
<td>Verse 2 softer than verse 1</td>
</tr>
<tr>
<td>6</td>
<td>MSopr</td>
<td>Ab4</td>
<td>34</td>
<td>1936</td>
<td>Germ</td>
<td>Piano</td>
<td>3.00</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Alto</td>
<td>G4</td>
<td>38</td>
<td>1971</td>
<td>Germ</td>
<td>Piano</td>
<td>2.33</td>
<td>3</td>
<td>Verse 2 and 3 much softer than verse 1</td>
</tr>
<tr>
<td>8</td>
<td>Tenor</td>
<td>C4</td>
<td>38</td>
<td>1979</td>
<td>Germ</td>
<td>Orchestra</td>
<td>2.10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Tenor</td>
<td>Bb3</td>
<td>35</td>
<td>1968</td>
<td>Latin</td>
<td>Org&amp;Choir</td>
<td>2.13</td>
<td>2</td>
<td>Different choir arrangements in verses 1 and 2</td>
</tr>
<tr>
<td>10</td>
<td>Tenor</td>
<td>Bb3</td>
<td>52</td>
<td>1987</td>
<td>Latin</td>
<td>Piano</td>
<td>2.53</td>
<td>2</td>
<td>Same text in verse 1 and 2, verse 2 softer</td>
</tr>
</tbody>
</table>

All singers do not invariably stay on pitch, not even in phonogram recordings. The intonation was assessed in a listening test using seven professional musicians as subjects. They represented different professions where judging intonation was an essential part of their expertise, e.g., coach, bowed instrument player, choral director, recording producer, singing teacher. The subjects were given a cassette recording of all ten recordings plus a form, where scores of Ave Maria melody line was reproduced on ten sheets, one for each recording. The subjects were asked to mark by arrows and/or circles all tones that they perceived as being out-of-tune as compared with the tuning of the accompaniment. In the test tape, all recordings appeared only once, without replications, since it was assumed that these experts would recognise and remember each of the recordings. This assumption was confirmed by the fact that, indeed, one subject correctly guessed the names of nearly all singers. The subjects were free to listen as many times as they wanted to each recording.

Nine of the ten recordings contained two verses of the song, each starting and ending with the text “Ave Maria”. Replicability of intonation patterns was analysed by
Results

Accuracy of F0 measurement
The accuracy of the F0 data obtained by Sonagraph measurement were checked. A professional mezzo-soprano was asked to sing the song with piano accompaniment. Her singing was recorded from an audio microphone and an electroglottograph producing a signal reflecting the vocal fold contact area. This recording was analysed in two ways. The audio recording was analysed by the Sonagraph procedure described above. The electroglottograph recording was analysed by means of the SWELL pitch tracking program using the double peak picking strategy (Ternström, 1991). The agreement between the two methods was quite high, displaying a maximum error of 12 cents for one single tone. The mean discrepancy was 2.5 cent, SD 2.7 cent. This indicates that the Sonagraph method was reliable.

Listening test
The results from the listening test did not show a great deal of agreement. Astonishingly few tones escaped complaints about tuning by all listeners. The number of complaints received from the different listeners varied greatly, between 15 and 100. With regard to the decision whether the tone was flat or sharp, there was remarkably low degree of agreement. One of the listeners used only arrows pointing upward. These observations suggest that it is easier to decide whether a tone is out-of-tune or not, than to decide if it is sharp or flat.

Replicability of F0
As mentioned, replicability of intonation was analysed by comparing the first and last three measured tones of the first verse with their replications in the second verse. The results are shown in Figure 2 in terms of the F0 curves for singers 1, 2, 3, and 4. There are many cases of a striking similariy in the deviations from the solid reference line representing ETT. For instance, tones 3 and 25 were both clearly lower than tones 2 and 24 in many singers.

In Figure 3, all singers’ analysed replicated tones can be studied in terms of the mean deviation from the ETT over the entire duration of the tone. Disregarding single discrepancies of up to 35 cent, the consistency was mostly quite high. The mean across all singers of the absolute difference between the same tones in verse 1 and 2 amounted to 8.0 cent (SD=7.1 cent). Singer 10 used the same text in both verse 1 and 2, performing the second verse much softer than the first. His intonation tended to be clearly flatter in verse 2, the mean difference being 20.4 cent. For the remaining subjects, the mean intonation difference between verse 1 and 2 was no more than 6.4 cent (SD=4.7 cent). An ANOVA test was carried out for all singers for which all tones could be measured in both verses. The factors were singers x 6 tones x 2 verses. Main
effects of singer and tone were found significant ($p<0.01$), while main effect of verse was not significant ($p>0.05$). Thus, the singers tended to have similar intonation of these tones in verse 1 and 2.

Figure 2. Typical examples of the $F0$ curves observed for tones 1, 2, 3, 23, 24, and 25 in the performance of the first and second verse of the song (solid and dotted curves, respectively) as performed by singers 1, 2, 3, and 4. The thin straight line represents ETT.
**Intonation**

Figure 4 shows some examples of F0 curves from the material assorted according to the tone. Figure 4a shows all tones with zero complaints, and Figure 4b shows all tones with more than four complaints. With three exceptions all tones judged to be out-of-tune by at least five of the listeners were flat relative to the ETT, while several of the tones with zero complaints showed a rather close agreement with the ETT. However, some tones clearly deviated from ETT. Tones 3 and 25 were clearly below ETT, while some examples of tones 4, 19 and 21 were above. This suggests that the perception of out-of-tune does not correspond to agreement with ETT. Contextual effects seem influential.

Three singers showed a constant tendency to flat intonation relative to the ETT of the accompaniment. In these cases, the first tones of the piece received many complaints about intonation while the number of complaints for flat tones occurring later in the song were comparatively few. This suggested that in these cases the tuning of the accompaniment was not used as the reference by the panel, but rather the pitch of the preceding long tone. To test this assumption, the data from these singers were related to the tuning of the preceding long tone throughout the piece, thus using a
Figure 4. Examples of F0 curves. Figure 4a shows all tones with zero complaints, and Figure 4b shows all tones with more than four complaints. The thin straight line represents ETT. The numbers left and right of the semicolon signs show the tone's mean deviation from ETT and the number of experts who found that the tone was out-of-tune.
running reference rather than the constant accompaniment reference. However, several instances of minor departures from ETT with many complaints occurred also when this reference was applied, although a better agreement between number of complaints and deviations from targets were observed in some cases. This indicates that the previous long tone may have been used as the basis of out-of-tune judgements only occasionally.

The effects of the musical context on the perception of being out-of-tune can be examined in Figure 5. Filled circles represent deviations from ETT of all tones which received zero complaints in the listening test and open circles represent tones which received more than four complaints. Interesting observations can be made. Most of the tones with zero complaints agreed within about ±7 cent with ETT. However, tones 3, 16, and 23 were clearly flat and tone 19 was clearly sharp as compared with ETT. The values for the tones with many complaints all lie outside of the areas occupied by the tones with no complaints. This suggests that the marked areas represent the F0 demand for a correct intonation. For tones 3 and 22, values very close to ETT were perceived as out-of-tune, while values further away from ETT were perceived as in-tune. Thus, for some tones perfect agreement with ETT was not the criterion for correct intonation. In other words, the listeners did not use disagreement with the tuning of the accompaniment as the criterion for their complaints. It is also interesting that the span of the tones perceived as in-tune varied within about ±7 cents for all tones except tones 6 and 7, where the variation was much greater. This supports the conclusion that singers mostly need to match the target pitch with an accuracy of about ±7 cent.

![Figure 5](image)

**Figure 5.** Deviations from ETT, averaged over the entire duration of the tones, for tones perceived as out-of-tune by none of the listeners (filled circles) and by more than four of the seven listeners (open circles).
Certain principles can be identified which seem to affect tuning. One was that high tones are sharpened and low tones are flattened. This tendency is illustrated in Figure 6, showing the mean deviation from ETT across all singers for all 25 tones. In the graph, the deviations are plotted as function of the pitch of the tone. The correlation is significant thus corroborating a rule previously found in analysis by synthesis of musical performance (Friberg, 1991).

Some results support another performance rule originally emerging from the same work, "melodic intonation" (Friberg, 1991). This rule states that scale tones on the dominant side of the circle of fifths are played sharp, while those on the subdominant side are played flat. The deviations are greater for tones far down the circle than for tones higher up. The scale tone at "noon" is the root of the prevailing chord. The purpose of the rule seems to be to enhance the differences between the various scale tones. Five of the singers (1, 2, 4, 6, and 8) showed significant linear correlation between their deviations from ETT and the deviations predicted by this rule.

Discussion

One interesting finding is that agreement with ETT is not a reliable correlate of the perception that a tone is in-tune. The departures for the tones, which were perceived as in-tune by all listeners, in some cases amounted to about +10 cent or -10 cent.
depending on the musical context. This amount may seem small, but is slightly greater than the singers' replicability of intonation. Thus, these deviation do not seem to reflect the limitations of the singers' pitch control. Rather they would be intended. There seems good reasons to assume that singers and listeners agree on certain modifications of the ETT recipe.

The results also give a hint as to what principles underlie these expected departures from the ETT. One principle was to sing high tones sharper and low pitches flatter than ETT. The mean stretch across singers amounted to 25 cent per octave. This tendency to stretch intonation with pitch has been observed also in instrument performances (Sundberg, 1982). It is also present in piano tuning, although it occurs for purely acoustical reasons in this case and amounts to only about a tenth of the singers' deviations (Shuck & Young, 1943). Thus, it is unlikely that piano tuning is the explanation of these deviations from ETT. Rather they seem to respond to a general desire in music listening.

Another principle is the tendency to deviate depending on the scale tone. Also in this case, parallel findings have been made in analyses of instrument performances (Sundberg, 1995). Professional players of bowed instruments all agree that e. g. the scale tone G# must be played sharper than the scale tone Ab, even though they are represented by the same key on keyboard instruments. It is not surprising that singers adhere to the same principle.

There must be other principles than these two. While the tones that received no complaints on intonation generally were found within a band of no more than 15 cent or less, much greater variations were observed on some tones. For example, those representations of tone 7 in the piece, which received no complaints, varied within no less than 55 cent. This tone is musically stressed, initiating a group of three long notes which form a descending chromatic sequence. It is also the highest one of all long tones in the piece. It is possible that this musical context offers the singer a great intonation liberty. This assumption is supported by the fact that none of the various intonations of this tone received complaints from a majority of the listeners. The great variation of the intonation of this particular tone suggests that intonation is used as an expressive mean in sung performance. It further shows that the above mentioned intonation rules are not compulsory.

Perhaps, the most remarkable result is that, on average, most singers replicated their deviations from ETT within ±6 cent for the same tone in the two verses. This is an astonishingly high degree of accuracy, given the fact that the vibrato extent typically is ±70 cent in these recordings, i.e., more than ten times wider. Moreover, it is very close to the DL of perceived pitch of sine tones, amounting approximately to 5 cent (0.2%) (Moore, 1989).

It is interesting to view this accuracy in the light of the pitch regulating system of the human voice. For example, fundamental frequency of phonation is affected by subglottal pressure. A change of one cm H2O subglottal pressure produces a pitch shift on the order of magnitude of 4 Hz, i.e., 16 cent at 440 Hz and 32 cent at 220 Hz. Assuming that singers use a subglottal pressure in the vicinity of 10 cm H2O this implies that they need to realise the target pressures within ±0.5% at 440 Hz and
+0.25% at 220 Hz. This would be a remarkable accuracy of subglottal pressure control. However, we do not know to what extent it is really achieved. Our results merely show that top class singers manage to control the summed contributions of all factors affecting pitch such that the errors are close to the capability of human pitch discrimination.

Intonation may possibly be used by singers as an expressive means, and the expression may not necessarily be identical between verses. Therefore, replicability could not be analysed by simply comparing the same tone in the different verses.

Conclusions

This study has revealed that an astonishingly high number of tones in first rate performances are perceived as out-of-tune by expert listeners. When singers sing the same tone in different verses of a song they are capable of replicating mean fundamental frequency within 6 cent, or 0.5% of the fundamental frequency. Singers mostly seem to need to match the target fundamental frequency within a band of about 7 cents, but musical expression may apparently widen this band considerably in some musical situations. Thus, deviation from the ETT is not the sole correlate of the perception of out-of-tune.

Acknowledgements

The data presented in this investigation were collected by co-author EP with assistance of Olof Essle and co-author JI, who also ran the listening test. The kind and competent assistance of Lars Frydén in the determination of accompaniment tuning and of Anders Friberg in advice on statistical calculations are gratefully acknowledged. The work was partly supported by a grant from the Swedish Council of Technical Sciences.

References


