Elsevier Editorial System(tm) for Journal of Voice
Manuscript Draft

Manuscript Number:

Title: Intonation and Expressivity A single case study of Classical Western Singing

Article Type: Full Length Article

Keywords: Intonation; Sharpening; Expressivity; Western Classical Singing; Phase-peak Tones

Corresponding Author: Dr. Johan Sundberg, PhD.

Corresponding Author's Institution: KTH Voice Research Centre

First Author: Johan Sundberg, PhD.

Order of Authors: Johan Sundberg, PhD.; Johan Sundberg, PhD; Filipa M Lã, PhD

Abstract: Previous studies have shown that singers tend to sharpen phase-peak tones as compared with equally tempered tuning (ETT). Here we test the hypothesis that this can serve the purpose of musical expressivity. Data were drawn from earlier recordings, where a professional baritone sang excerpts as void of musical expression as he could (Neutral) and as expressive as in a concert (Concert). Fundamental frequency averaged over tones was examined, and compared with ETT. Phrase-peak tones were sharper in agitated examples, particularly in the Concert versions. These tones were flattened to ETT using the Melodyne software. The manipulated and original versions were presented pair-wise to a musician panel, asked to choose the more expressive version. By and large, the original versions were perceived as more expressive, thus supporting the common claim that intonation is a mean for adding expressivity to a performance

Cover Letter

2012-07-17

JOHAN SUNDBERG
PROFESSOR OF MUSIC ACOUSTICS (RETIRED)
DEPT OF SPEECH MUSIC HEARING
SCHOOL OF COMPUTER SCIENCE & COMMUNICATION KTH
SE-100 44 STOCKHOLM
PHN +468 790 7873, FAX +468 790 7854
MOBILE PHONE +46 70 740 78 73
email pjohan@speech.kth.se

Journal of Voice

Here we submit a manuscript on intonation and expressivity in singing, which has not been published elsewhere.

With best regards,

Johan Sundberg

*Manuscript

Intonation and Expressivity

A single case study of Classical Western Singing*

Authors:

Johan Sundberg

Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH, Stockholm, Sweden

Filipa M. B. Lã

Department of Communication and Arts, INET-MD, University of Aveiro, Portugal

Evangelos Himonides

Institute of Education, University of London

* This work, presented as a poster at the 40th Annual Symposium Care of the Professional Voice, Philadelphia PA, in Jun 2011, won the "Best Poster Prize".

Abstract

Previous studies have shown that singers tend to sharpen phase-peak tones as compared with equally tempered tuning (ETT). Here we test the hypothesis that this can serve the purpose of musical expressivity. Data were drawn from earlier recordings, where a professional baritone sang excerpts as void of musical expression as he could (Neutral) and as expressive as in a concert (Concert). Fundamental frequency averaged over tones was examined, and compared with ETT. Phrase-peak tones were sharper in agitated examples, particularly in the Concert versions. These tones were flattened to ETT using the Melodyne software. The manipulated and original versions were presented pair-wise to a musician panel, asked to choose the more expressive version. By and large, the original versions were perceived as more expressive, thus supporting the common claim that intonation is a mean for adding expressivity to a performance.

Keywords: Intonation; Sharpening; Expressivity; Western Classical Singing; Phase-peak Tones

Introduction

Intonation is commonly regarded as an important aspect of music performance. Already in the 1930s, Seashore and associates (1938) measured fundamental frequency, F0, in recordings of renowned singers and revealed considerable departures from equally tempered tuning, henceforth ETT (Seashore, 1938). He observed that ”*long notes tend to begin slightly flat and are gradually corrected, as if by hearing*” (Seashore, 1938, p.265). Although he termed the frequencies of the ETT the “*true*” values, he also assumed that these deviations have an “*aesthetic value in phrasing and harmonic balancing*” (Seashore, 1938, p.267).

Rapoport (1996) further developed this idea using FFT analysis on selected tones sung in commercial recordings by famous vocal artists. He found momentary deviations from the mean F0 during a tone, “*microintonation*”, and classified them into a system of “*singing mode categories*”. He hypothesized that “*the singer uses microintonation in a very methodical way as his or her means of expressiveness beyond the written score*” (Rapoport, 1996, p. 112-113).

Singers’ F0 accuracy was further explored in a study of ten renowned singers’ commercial recordings of the song *Ave Maria* by Franz Schubert (Prame, 1997; Sundberg et al., 1996). The mean F0 values of tones were compared with the ETT of the accompaniment. The results revealed deviations from ETT, henceforth ΔETT, of about ±40 cent, thus exceeding almost by an order of magnitude the just-noticeable difference for frequency discrimination. An expert panel was asked to identify all tones that sounded out of tune in this material. The result showed that: (1) for most tones, the tolerance zone for “*in tune*” was about ±10 cent, i.e., tones outside this zone tended to be heard as out of tune; and (2) the mean F0 value of certain tones, perceived as in tune, could deviate from ETT by as much as +35 cent or -20 cent, thus widely exceeding the tolerance zone (Sundberg et al., 1996).

Other commercial recordings of premier singers' performances were recently analyzed in some detail (Sundberg, 2011a). Figure 1 shows an example taken from the first Kyrie movement of Verdi's *Messa da Requiem* as performed by the late tenor *Jussi Björling*. The singer started the long, top note of the phrase, henceforth phrase-peak note, at ETT, but gradually sharpened it, ending at more than +50 cent.

Please insert Figure 1 about here

Further measurements of other world famous tenors' intonation showed that such sharpening, by no means, is specific to Björling (Sundberg, 2011b). For example, sharp intonation of a phrase-peak tone was noted also in *Carlo Bergonzi's, Nicolai Gedda's* and *Luciano Pavarotti's* performances of an excerpt from Verdi's opera *Aida*. The result was a descending octave interval that exceeded a 2:1 ratio. Such octave stretch has been found to be preferred by musically trained listeners (Sundberg, 1982; Sundberg and Lindqvist, 1973).

Of course, ΔETT in professional music performances are not specific to singers (for reviews, see e.g., Gabrielson, 1999; Morrison and Fyk, 2002). Also, in the *Director Musices* system of musical performance rules, derived from analysis-by-synthesis, one of the formulated principles is to sharpen contrasts, and this can concern not only durational patterns but also sharpening of the pitch (Friberg, 1995).

While some of singers' ΔETT might be caused by timbral effects (Vurma et al., 2010), most authors assume that such deviations are performed for the sake of expressiveness (see e.g. Fyk, 1995). However, this assumption has not yet been formally tested. The aim of the present study was to find out to what extent intonation in singing can be perceived as contributing to expressivity. Two experiments were carried out: in the first, F0 was extracted for all tones in all examples and the averages were compared to ETT; in the second, a listening test was carried out, presenting paired stimuli with intonation differences to a panel of music experts.

Data selection

The material was collected from an earlier study on expressivity in singing (Sundberg, Iwarsson and Hagegård, 1995). Opera baritone *Håkan Hagegård* was recorded while performing a set of song excerpts in two different ways: (1) as void of musical expression as he could (Neutral version), and (2) as in a public performance (Concert version). The excerpts are listed in Table 1.

Please insert Table 1 about here

These recordings were presented to two expert panels. The task for the first panel was to identify the pairs that differed most with respect to expressivity. The excerpts that were perceived as most differing were selected for the test with the second panel. The aim here was to identify the prevailing emotional colours of the performed excerpts. The alternatives were Angry, Happy, Loving, Hateful, Sad, Solemn and Secure. Five excerpts were perceived as Loving, Sad or Secure, thus sharing a peaceful character, while six perceived as Angry, Solemn or Happy were assumed to share an excited, agitated mood.

Acoustic analysis of the Neutral and Concert versions revealed a number of differences, both in the agitated and in the peaceful examples: loudness, tempo, phonation type, and rate of change of sound level. However, F0 was not examined. Therefore, in the first experiment of the present study F0 was measured in the agitated and peaceful examples, and compared between the Neutral and Concert versions.

Experiment 1 - F0 extraction

Method

The Corr module of the Soundswell signal workstation was used for F0 extraction. This module applies an autocorrelation strategy and shows in terms of a gray scale the correlation between the signal in two adjacent analysis windows (Granqvist and Hammarberg, 2003). The output is an *smp* file containing the F0 curve.

Using the Histogram module of the Soundswell software, the average F0 value was determined in all examples for all tones that had a duration exceeding one complete vibrato period (about 170 ms). Each measurement included a sequence of complete vibrato periods. The values obtained were then expressed in semitones using 440 Hz for the pitch of A4 as the reference. However, it turned out that the singer often deviated slightly from this reference. Therefore, the frequency that yielded a mean ΔETT of zero cent across the entire excerpt was used as a new reference for ETT. For example, if $F0_{A4}$=440 Hz yielded a mean ΔETT of +15 cent, the 440 Hz reference was replaced by a 15 cent sharper reference (i.e. 443,83 Hz). This reference was used for measuring the average ΔETT also in the Neutral version.

Results

Figure 2 compares, in terms of a scatter plot, the ΔETT for all tones measured in the Neutral ($\Delta ETT_{Neutral}$) and in the Concert ($\Delta ETT_{Concert}$) versions. $\Delta ETT_{Concert}$ ranged between +56 and -76 cent and $\Delta ETT_{Neutral}$ between +53 and -77 cent. The trend line (R^2 = 0.106) for the data points was

$$\Delta ETT_{Neutral} = 0.3 * \Delta ETT_{Concert} - 4.3 \text{ [cent]}$$

Thus, on average, the $\Delta ETT_{Concert}$ tended to be considerably greater than $\Delta ETT_{Neutral}$.

Please insert Figure 2 about here

A paired samples t-test was used to assess which of the Concert and Neutral versions showed a significant difference in ΔETT. The Falstaff example was the only one for which a significant difference was found ($t_{(15)} = 4.781$; $p < 0.05$).

A closer examination of the phrase-peak tones revealed a systematic trend. Table 2 lists their durations and ΔETT values in the two versions. While the tone durations did not differ significantly, systematic differences were observed for ΔETT.

Please insert Table 2 about here

Given these results, two comparisons seemed relevant for each of the examples: (i) between agitated and peaceful examples; and (ii) between the pairs of Concert and Neutral versions.

Table 3 lists the results of the ΔETT for the agitated and peaceful examples. Smaller ΔETT values were found for the peaceful examples (Figure 3): mean $\Delta ETT_{Agitated}$ = 29.1 cent; SD = 29.4 cent; mean $\Delta ETT_{Peaceful}$ = 13.1 cent; SD = 27.6

cent; however, a paired samples T-test showed that both these differences failed to reach significance.

Please insert Table 3 about here

With respect to the Concert and Neutral comparison (Figure 3), ΔETT values were significantly greater for Concert versions of the agitated examples only (one sample T-test: $t_{(9)} = 2.94$; $p = 0.017$). This seemed to support the hypothesis that these deviations were related to the expressivity of the performance.

Please insert Figure 3 about here

Experiment 2 - Listening test

Method

To test the last mentioned hypothesis, a listening experiment was carried out. The strategy was to explore the perceptual effect of artificially flattening, to ETT, the sharp phrase-peak tones in the five agitated examples. The flattening was achieved by means of the *Melodyne* software, which allows changing of F0 in audio recordings (http://en.wikiaudio.org/Melodyne: Tools, accessed 20 January 2012).

Special care was taken to find out to what extent *Melodyne* affected not only F0 but also other acoustic parameters. This was tested in the following way. A vowel was synthesized on the *Madde* software (by Svante Granqvist); its F0, formant frequencies and bandwidths are listed in Table 4.

Please insert Table 4 about here

The synthesized vowel was then manipulated using *Melodyne*. Two versions were thus obtained, the *original* and the *manipulated*. In the latter, F0 (110 Hz) was lowered by 6 Hz (or 80 cent). The formant frequencies and bandwidths of these versions were analyzed by inverse filtering, using the *Decap* custom made software (by Svante Granqvist), described elsewhere (Sundberg et al., 2011). The inverse filtering analysis returned the original settings within an error that ranged from -37 to +14 cent (Table 4). However, *Melodyne* was found to shift not only F0, but also the formants; F0 shift was measured as -97 cent (rather than 80 cent), and F1 and F2 were shifted by a similar amount (89 and 88 cent). Such formant shifts would have an effect on voice timbre. In fact, this is also implicitly acknowledged in the *Melodyne* manual (http://en.wikiaudio.org/Melodyne: Tools), although with reference to the effect of an increase rather than a decrease of F0: "*Sometimes in Melodyne when you move a note, say for example to a higher pitch, it may begin to sound 'chipmonk-ish'. Moving the formant back down a little while keep the note in its new position can alleviate this unwanted characteristic*." Such a "*chipmonkish"* timbre would be associated with an upward shift of the formant frequencies.

Please insert Table 4 about here

Melodyne was then used for flattening the F0 of the phrase-peak tones, such that they matched those of ETT. The amount of flattening used for the different agitated examples were listed in Table 3.

The original and manipulated versions of the five excerpts were then arranged pair-wise in a listening test, where experts were asked to judge which version in each pair sounded more expressive.

A preliminary version of this test was piloted with a group of ten musicians. It contained a total of 20 presentations of paired stimuli, each of the five stimulus pairs being presented four times: two with the Neutral and two with the Concert versions as the first in the pair.

The listeners were given a written instruction on a response sheet, asking them to decide if they found the first or the second version in the pair more expressive: "*From the paired examples following presented, please tick the one that you perceive as most expressive*". The feedback from these listeners revealed that they found it extremely difficult to determine which version was more expressive.

In the subsequent final version of the test, 42 expert listeners, music students or teachers, were asked to pay special attention to the peak-phrase tone, which was given in the response sheet in terms of the corresponding word in the lyrics. For each stimulus pair there were two boxes in the response sheet, and the listeners gave their responses by ticking the box corresponding to the version that they perceived as more expressive. No scores were given. The same stimuli order was used for all. At the end of the response form, an open question was included asking the listeners what characteristic they found most relevant to their choices.

Results

As the data were dichotomous nominal the results of the listening test were first evaluated by means of one-sample binominal test (Table 5). The underlying assumption was that the proportion of responses was the same for the original and manipulated versions, i.e. that the proportion would not deviate significantly from 0.5.

Please insert Table 5 about here

Significant deviations were found in all examples, except for “Sie(ist dein)”, see Figure 4. However, the order of the versions within the pair seemed relevant to the responses. When the original appeared as the second in the pair, it was consistently rated as being more expressive for “Cor” and ”Hei(lig)”, for both presentations of these pairs. For “Me(sser)” significant preference for the original version was observed only the first time it appeared in the test. The trend to prefer the second version in the pair was so strong in the case of “Herz(ensnacht)” that the manipulated version was significantly preferred when it appeared as second in the pair while the original version was preferred both times it appeared as the second in the pair.

Please insert Figure 4 about here

To investigate whether there was an overall preference for the original version in the pairs, the proportions of votes for the original version were averaged (1) for all excerpts and (2) for each individual excerpt. A one sample T-test was applied to assess whether these means differed significantly from 0.5, at a confidence level of 95%. With respect to the average across all excerpts, significant proportion differences were found showing that the original version was perceived as being more expressive than the manipulated. With respect to individual excerpts, the original versions of the examples “Me(sser)”, ”Hei(lig)” and ”Cor” were rated as significantly more expressive. For examples “Herz(ensnacht)” and “Sie (ist dein)”, the preference for the original versions failed to reach significance. These results are listed in Table 6.

Please insert Table 6 about here

As mentioned, the listeners were given the opportunity to specify what determined their choice of version. Several responses mentioned the following: “support“, “energy“, “emphasis“, and “stress”. No responses mentioned intonation or timbral differences.

Discussion

Two main findings have emerged from our investigation: (1) the F0 measurements revealed that the singer sharpened phrase-peak tones in agitated but not in peaceful examples, and (2) the listening test demonstrated that expert listeners perceived this sharpening as adding to the expressivity. However, it is important to realize that the investigation may seem limited in certain respects.

Our study was based on recordings of a single singer, thus calling into question the general validity of the observations. On the other hand, the subject was an internationally touring opera and concert soloist. Further, as mentioned, the trend to sharpen phrase-peak tones has been observed also in commercial recordings of other internationally renowned singers (Sundberg, 2011a; Sundberg, 2011b). This supports the assumption that our results possess some general validity.

We flattened the phrase-peak tones in the Concert versions of the excerpts. An alternative way to test of the assumption that sharpening of phrase-peak tones adds to the expressiveness of agitated examples would have been to raise F0 of these tones in the Neutral versions. However, also the Neutral versions showed a sharpening of phrase-peak tones, even though smaller. Hence, the effect of the sharpening would then probably have been too subtle to produce a noticeable difference between the versions.

A crucial question for the validity of the listening test obviously is whether *Melodyne* introduced not only the transposition of F0 but also other audible effects. Our analyses revealed that for an F0 shift of 6 Hz, the software shifted also F1 and F2 by 40 or 50Hz. This is larger than the just noticeable difference for F1 and F2, which has been reported to vary between 10 and 30 Hz (Stevens, 1998). Hence, the manipulated versions may have been perceived as differing from the original ones with respect to timbre. On the other hand, our expert panel was asked to select the more expressive version in each stimulus pair; a timbral difference caused by a lowering of F1 and F2 is not very likely to be perceived as an expressivity increase of agitated examples. Also, no listener mentioned timbral differences as the characteristic on which they based their choice of version.

Expressiveness is generally regarded as a most important, though subjective, quality of performed music. Our panel consisted of qualified musicians, so it is reasonable to expect that they were skilled in recognising expressivity. Also, the risk of receiving random responses in a listening test was substantially reduced: (i) the test was previously piloted; (ii) the listeners' attention was cued to the specific word in the lyrics where the phrase-peak tone appeared; and (iii) these tones were presented in a musical context.

To our knowledge, this is the first time that expressive effects of intonation in singing are formally analysed by means of a perceptual evaluation by expert musicians. The results support the frequently made assumption that intonation is used as an expressive mean in music performance (see e.g. Morrison and Fyk, 2002). It is noteworthy that all phrase-peak tones in the agitated examples were sharpened, particularly in the Concert versions. This seems to be in accordance with previous measurements of musicians' intonation (Seashore, 1938). Incidentally, it also seems to agree with the saying sometimes heard among musicians "Better being sharp than out-of-tune".

If sharpening of phrase-peak tones constitutes a means to increase expressiveness, one might ask why this particular acoustic code is used for this purpose. According to the *Director Musices* grammar of music performance, a general principle underlying musicians' tuning strategies is to sharpen high pitched tones and flatten low pitched tones (Friberg, 1995). Although the sharpenings produced by that grammar are much smaller than those observed in the present study, our findings may be regarded as a qualitative corroboration of both these rules. It may also be relevant that elevated speaking F0 is a typical sign of agitation. According to Fyk (1995), players of bowed instrument tend to expand ascending intervals, and all phrase-peak tones are obviously preceded by ascending intervals.

Seashore (1938) noted that the mean F0 in singing is often flattened or sharpened, particularly for short tones, but did not specify how much. In the investigation of ten singers' intonations of long tones in Schubert's *Ave Maria*, it was found that expert listeners accepted some tones as "in tune", although they departed from ETT by +35 to -20 cent (Sundberg et al., 1995). The greatest ΔETT value, 82 cent was found in "Herz(ensnacht)", which, however, concerned a short tone, no more than 259 ms long. Perhaps duration is a relevant factor in determining the boundaries of "out-of-tune". This possibility may belong to the explanation why significant response differences between original and manipulated versions were obtained for only four of the five examples.

There was a strong trend in the panel to perceive the original version as more expressive when it appeared as the second stimulus in the pair. A possible explanation could be that the first version that appeared in the pair was compared to the subject's internal tuning reference, while when it appeared as the second in the pair it was compared to the first version that immediately preceded it. In any event, expectations play an important role in music listening (see, e.g. Huron, 2006).

When singing is perceived as out-of-tune, sample-accurate pitch manipulation software, such as Celemony Melodyne, Antares Autotune, Roland V-Vocal, are often used either to purely control the final vocal output for accuracy in intonation, or, perhaps, in order to evoke novel aesthetic experiences. Therefore, it is risky to resort to commercial recordings when investigating intonation in music performance, since the F0 parameter may have been manipulated.

Conclusion

Our investigation has shown that a professional singer may sharpen phrase-peak tones, sometimes even more than 50 cent, an observation in agreement with previous research. Most expert listeners did not notice this intonation strategy as a pitch effect when they compared the examples with versions, in which the intonation had been flattened to ETT. We also found that such sharpening may contribute to expressiveness; the singer sharpened tones more when he sang the examples as expressive as in a concert than when he sang them as void of expressivity as he could. Also, expert listeners perceived original versions as more expressive than those in which the intonation was in accordance with ETT. This observation provides scientific support for the commonly made claim that intonation can be used as an expressive mean in singing.

Acknowledgements

The authors acknowledge their great gratitude to *Håkan Hagegård,* who brilliantly provided the data analyzed in the present study. The authors are also indebted to the musicians who participated in the listening tests.

References

Friberg A. (1995) A Quantitative Rule System for Musical Expression, Music Acoustics, KTH Royal Institute of Technology, Stockholm.

Fyk J. (1995) Melodic intonation, psychoacoustics, and the violin Organ Publishing House, Zielona Góra.

Gabrielson A. (1999) The Performance of Music, in: D. Deutsch (Ed.), The Psychology of Music, Academic Press San Diego. pp. 501-602.

Granqvist S., Hammarberg B. (2003) The correlogram: A visual display of periodicity. Journal of the Acoustical Society of America 114:2934 - 2945.

Huron D. (2006) Sweet anticipation: music and the psychology of expectation MIT Press, Cambridge.

Morrison S.J., Fyk J. (2002) Intonation, in: R. Parncutt and G. E. McPherson (Eds.), The Science and Psychology of Music Performance, Oxford University Press, Oxford. pp. 183-197.

Prame E. (1997) Vibrato extent and intonation in professional Western lyric singing. Journal of Acoustical Society of America 102:616-621.

Seashore C.E. (1938) Psychology of Music Dover Publications, Inc., New York.

Stevens K.N. (1998) Acoustic Phonetics The MIT Press, Cambride, Massachusetts.

Sundberg J. (1982) In tune or not? A study of fundamental frequency in music practise. STL-QPSR 23:49-78.

Sundberg J. (2011a) Some observations on operatic singers' intonation. Interdisciplinary Studies in Musicology 10:47-60.

Sundberg J. (2011b) On the Expressive Code in Music Performance, in: P. Lang (Ed.), Klangfarbe: Vergleichend-systematische und musikhistorische Perspektiven, Internationaler Verlag der Wissenschaften, Frankfurt am Main. pp. 113-127.

Sundberg J., Iwarsson J., Hagegård J. (1995) A singer's expression of emotions in sung performance, in: O. Fujimura and M. Hirano (Eds.), Vocal Fold Physiology: Voice Quality Control, Singular Press, San Diego.

Sundberg J., Prame E., Iwarsson J. (1996) Replicability and accuracy of pitch patterns in professional singers, in: P. Davis and N. Fletcher (Eds.), Vocal Fold Physiology, Controlling Complexity and Chaos, Singular Publising Group, San Diego. pp. 291-306.

Sundberg J.E.F., Lindqvist J. (1973) Musical octaves and pitch. Journal of Acoustical Society of America 54:922-929.

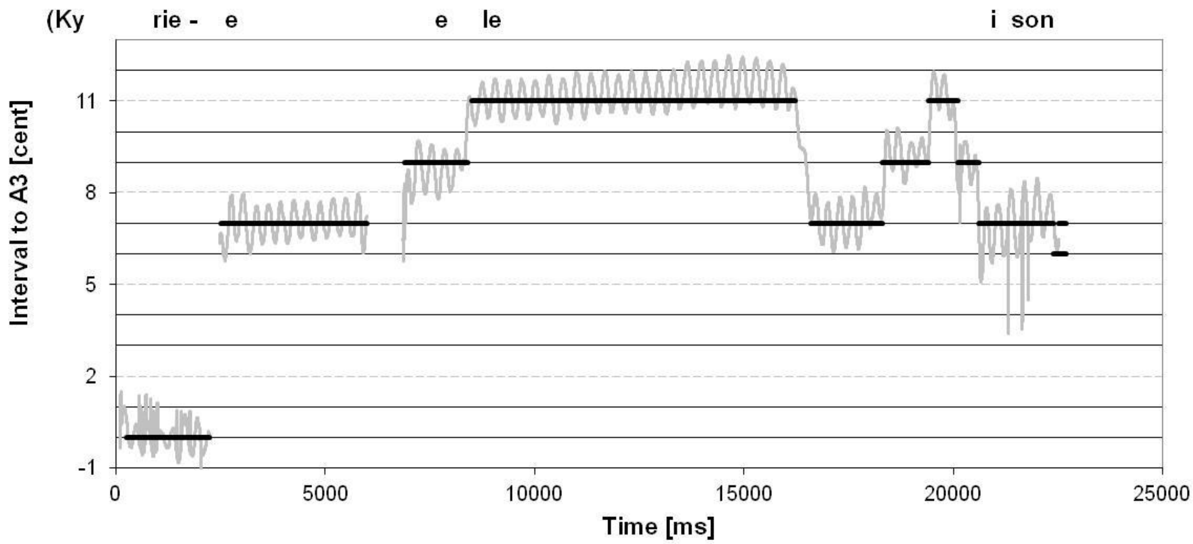
Vurma A., Raju M., Kuuda A. (2010) Does timbre affect pitch?: Estimations by musicians and non-musicians. Psychology of Music 39:291-306.

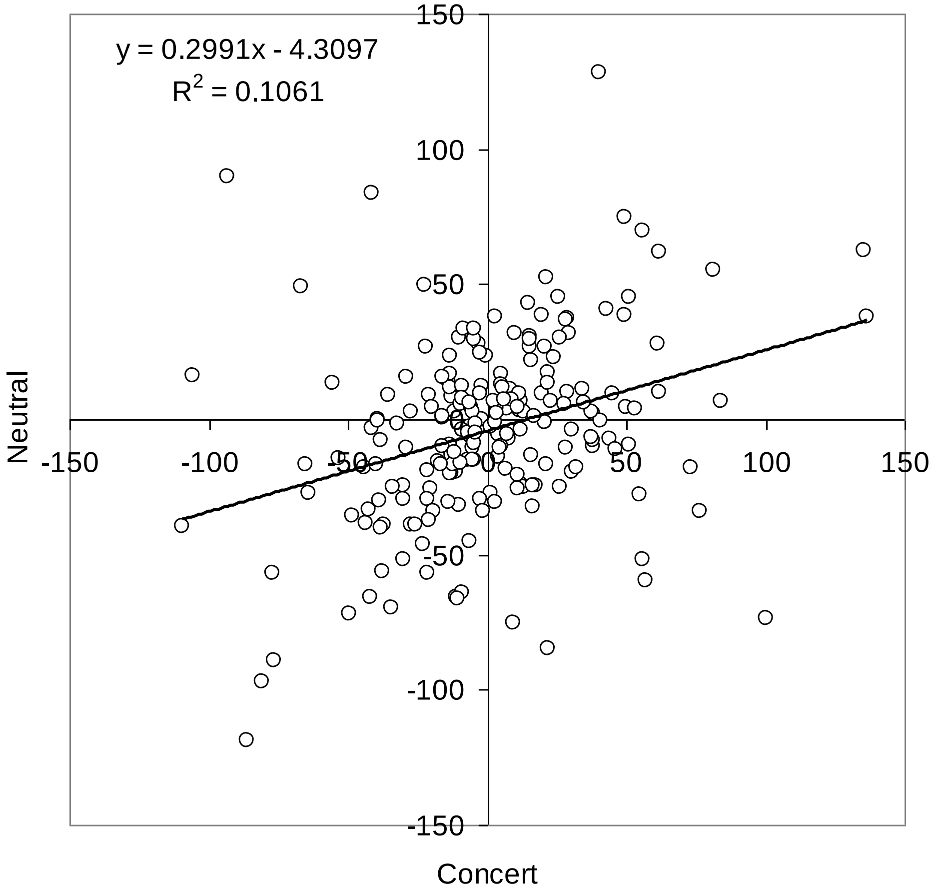
Figure 1. F0 pattern of Jussi Björling's rendering of the first Kyrie in Giuseppe Verdi's *Messa da Requiem* from the recording DECCA 467 119-2 made in 1960. The heavy horizontal lines show the F0 values according to the ETT.

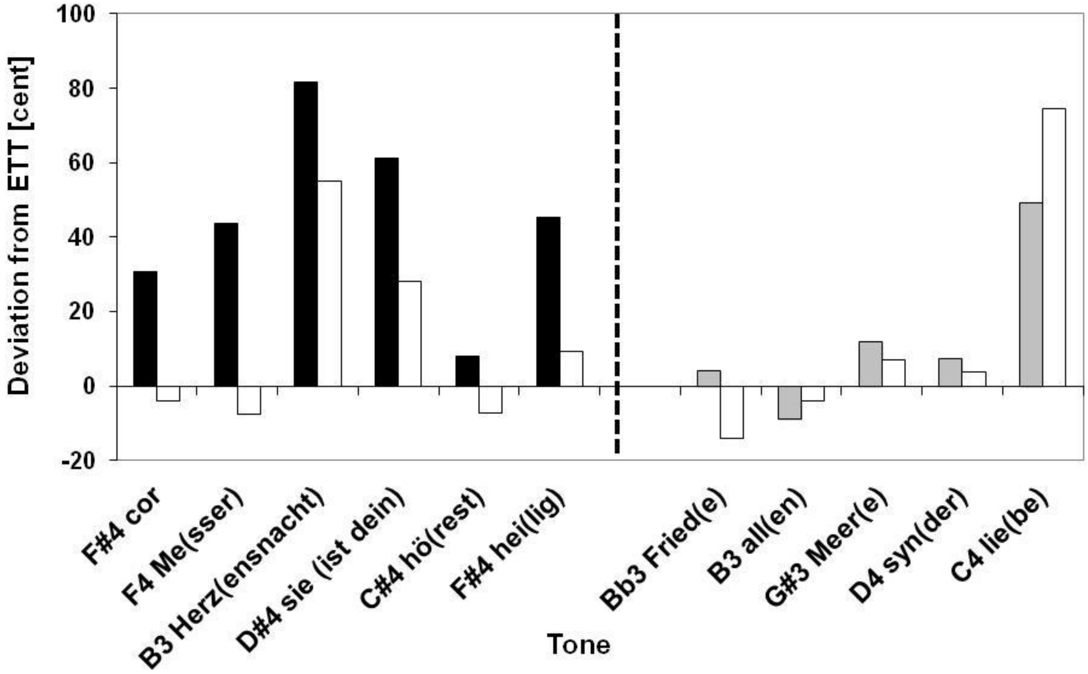
Figure 2. Comparison of the ΔETT for all tones measured in the Neutral and in the Concert versions. The equation refers to the trend line.

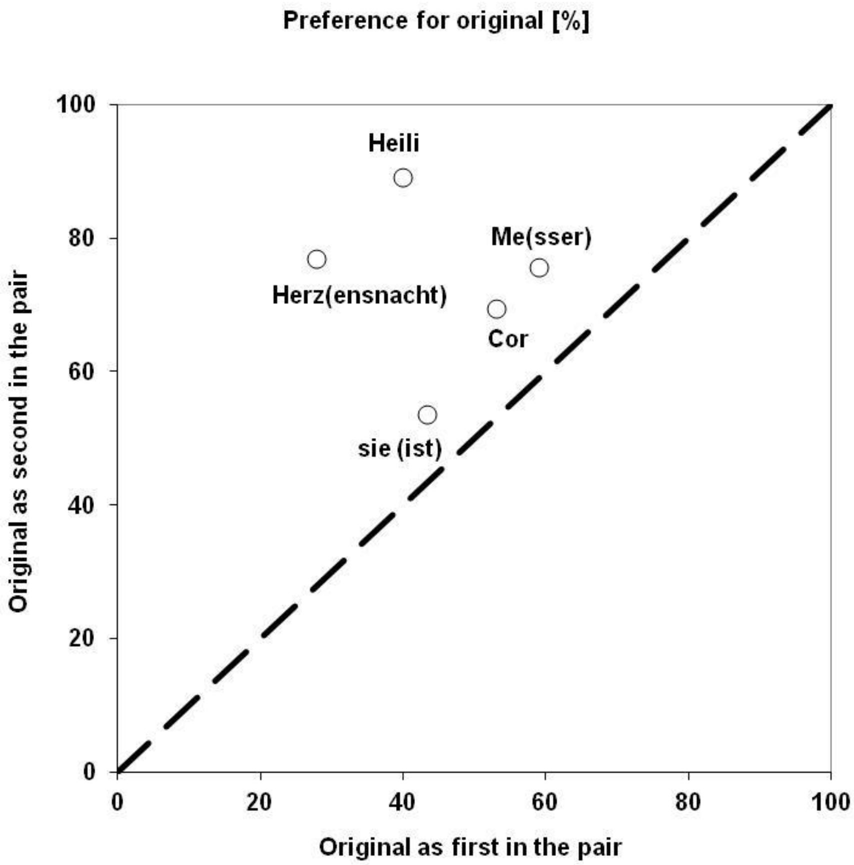
Figure 3. ΔETT of the mean F0 of phrase-peak tones in the Concert and Neutral versions. Black and gray columns refer to agitated and calm examples, respectively, and filled and open columns to Concert and Neutral versions, respectively.

Figure 4. Percentage of votes for the original version being more expressive, when it appeared as the first and the second stimulus in the pair.


(Ky
rie -
e
e
le
i
son
Interval to A3 [cent]
11
8
5
2
-1
0
5000
10000
15000
20000
25000
Time [ms]


y = 0.2991x - 4.3097
R2 = 0.1061
150
100
50
-50
-100
-150
-150
-100
-50
0
50
100
150
Neutral
Concert


Deviation from ETT [cent]
100
80
60
40
20
0
-20
F#4 cor
F4 Me(sser)
B3 Herz(ensnacht)
D#4 sie (ist dein)
C#4 hö(rest)
F#4 hei(lig)
Bb3 Fried(e)
B3 all(en)
G#3 Meer(e)
D4 syn(der)
C4 lie(be)
Tone


Preference for original [%]
Heili
Me(sser)
Herz(ensnacht)
Cor
sie (ist)
Original as second in the pair
Original as first in the pair
0
20
40
60
80
100

Table(s)

Table 1

Agitated and peaceful examples analysed in the previous investigation (Sundberg et al., 1996); b refers to bar number.

Agitated

Composer	**Piece**	**Section**	**Acronym**
G Mahler	Lieder eines fahrenden Gesellen, song # 3, b 5-11	Ich hab' ein glühend Messer...	Me(sser)
F Schubert	Erlkönig, b 72-79	Mein Vater, mein Vater…	Vater
R Schumann	Liederkreis XII , b 18-26	Und der Mond...	Sie(ist dei
R Schumann	Dichterliebe VII, b 12-18	Wie Du auch strahlst…	Herz(ensn
R Strauss	Zueignung, b 21-29	Und beschworst darin die Bösen...	He(ilig)
G Verdi	Falstaff, Ford's monologue, b 24-31	Laudata sempre sia...	Cor

Peaceful

F Schubert	Du bist die Ruh, b 8-15	Du bist die Ruh...	Frie(de)
F Schubert	Wanderers Nachtlied, b 3-14	Über allen Gipfeln ist Ruh…	All(e)
F Schubert	Nähe des Geliebten, b 3-8	Ich denke dein...	Mee(re)
R Schumann	Dichterliebe VI, b 31-42	Es schweben Blumen und Englein...	Lie(be)
F Mendelsohn	Paulus, Aria # 18, b 5-13	Gott sei mir gnädig...	Syn(der)

Table 2

Durations and mean deviation from ETT (ΔETT) of the phrase-peak tones in the agitated examples, for the Neutral and Concert versions

Example	**Specified word**	**Tone duration [s]**		**ΔETT [cent]**	
		Neutral	**Concert**	**Neutral**	**Concert**
Ford's monologue	**Cor**	2.01	1.90	-4	31
Zueignung	**Hei**(lig)	.92	1.55	9	45
Ich hab' ein glühend Messer	**Me**(sser)	.42	.34	-7	44
Wie Du auch strahlst	**Herz**(ensnacht)	.40	.25	51	82
Und der Mond	final **Sie** (ist dein)	.28	.46	28	61

Note. Bolded syllables refer to the phrase-peak tones analyzed

Table 3

F0 data for phrase-peak tones. "Word" lists the word in the lyrics on which the phrase-peak tone appeared. Columns "F0 in ETT" and "Mean F0" present the F0 values of these tones in ETT and in the singer's Concert version. Column ΔETT shows how much Mean F0 deviated from ETT, which in the case of the agitated examples equal the size of the F0 shift produced by the Melodyne manipulation.

	Word	Pitch	F0 in ETT [Hz]	Mean F0 [Hz]	SD [Hz]	ΔETT [cent]
Agitated	Cor	F#4	369.6	3762	13.4	31
	Me(sser)	F4	350.9	359.9	17.8	44
	Herz(ensnacht)	B3	245.6	257.4	13.8	82
	Sie (ist dein)	D#4	305.0	316.0	16.9	61
	Hö(rest)	C#4	268.3	269.6	14.0	8
	Hei(lig)	F#4	363.8	373.5	13.4	45
Peaceful	Frie(den)	Bb3	235.6	236.2	4.8	4
	All(e)	B	242.5	241.3	3.6	-9
	Mee(re)	G#3	205.5	206.9	5.0	12
	Syn(der)	D4	289.1	290.4	6.2	7
	Lie(be)	C4	257.8	265.2	7.5	49

Table 4

Effects of the Melodyne manipulation on formant frequencies (Fn) and bandwidths (Bn). Column "Used" lists the frequencies of Fn and their Bn used in the Madde synthesis. Columns "Measured" show the values obtained from the inverse filter analysis and columns "Difference" refer to the difference between the values listed in the "Used" and the "Measured" columns, for both Fn and Bn. The values obtained after the Melodyne F0 modification refer to those observed after the Madde synthesis had been modified by the Melodyne software.

	***Madde* synthesis**						**After Melodyne modification**				
	Used	***Measured***	**Difference**	***Used***	***Measured***	**Difference**	***Measured***	**Difference**		***Measured***	**Difference**
	Fn		Diff	Bn		Diff	Fn		Cent	Bn	
	[Hz]	[Hz]	[Hz]	[Hz]	[Hz]	[Hz]	[Hz]	[Hz]		[Hz]	[Hz]
F0	110	110	0	-	-	-	104	-6	-97	-	-
F1	700	709	9	47	28	-19	665	-35	-89	40	-7
F2	1050	1057	7	55	32	-23	998	-52	-88	38	-17
F3	2500	2496	-4	114	77	-37	2480	-20	-14	80	-34
F4	2700	2702	2	108	88	-20	2699	-1	-1	87	-21
F5	2900	2920	20	116	130	14	2900	0	0	122	6

Table(s)

Table 5

Results of one-sample binominal test, showing the votes for the original version being more expressive. From left to right, the columns represent: examples; where the stimulus pair appeared in the test (Appearance of pair in test); what version was first in the pair (Order in the pair); number of votes for the original as being more expressive (n), proportion of these votes (Proportion); and p values.

Example	Appearance in test	Order in the pair	n	Proportion [%]	p (2-tailed)
Cor	1st	Original first	21	51	1.000
		Manipulated first	28	68	.028*
	2nd	Original first	23	56	.533
		Manipulated first	29	71	.012*
He(ilig)	1st	Original first	15	37	.117
		Manipulated first	37	90	<.001*
	2nd	Original first	18	44	.533
		Manipulated first	36	88	<.001*
Me(sser)	1st	Original first	26	63	.117
		Manipulated first	35	85	<.001*
	2nd	Original first	23	56	.533
		Manipulated first	27	66	.06
Herz(ensnacht)	1st	Original first	11	27	.004*
		Manipulated first	34	83	<.001*
	2nd	Original first	12	29	.012*
		Manipulated first	29	71	.012*
Sie(ist dein)	1st	Original first	20	49	1
		Manipulated first	19	46	.755
	2nd	Original first	16	39	.211
		Manipulated first	25	61	.211

Note. Test value = 0.5; *significance level $p < 0.05$; ($n = 41$)

Table 6

Results of the one sample T-test (t) and associated p values for the five excerpts. SD refers to the standard deviations of the proportion of votes given to each of the five original versions.

Example	**Mean (*SD*)**	**One sample T-test**	
		t(40)	**p**
All examples	***.59 (.11)***	***4.99***	***<0.001***
Cor	.61 (.23)	3.04	<.001*
Hei(lig)	.63 (.23)	3.61	<.001*
Me(sser)	.67 (.21)	5.36	<.001*
Herz(ensnacht)	.52 (.18)	.85	.4
Sie (ist dein)	.48 (.25)	-.31	.76

Note. Test value = 0.5; *significance level p <0.05; (n = 41)

Figure(s)
Click here to download high resolution image

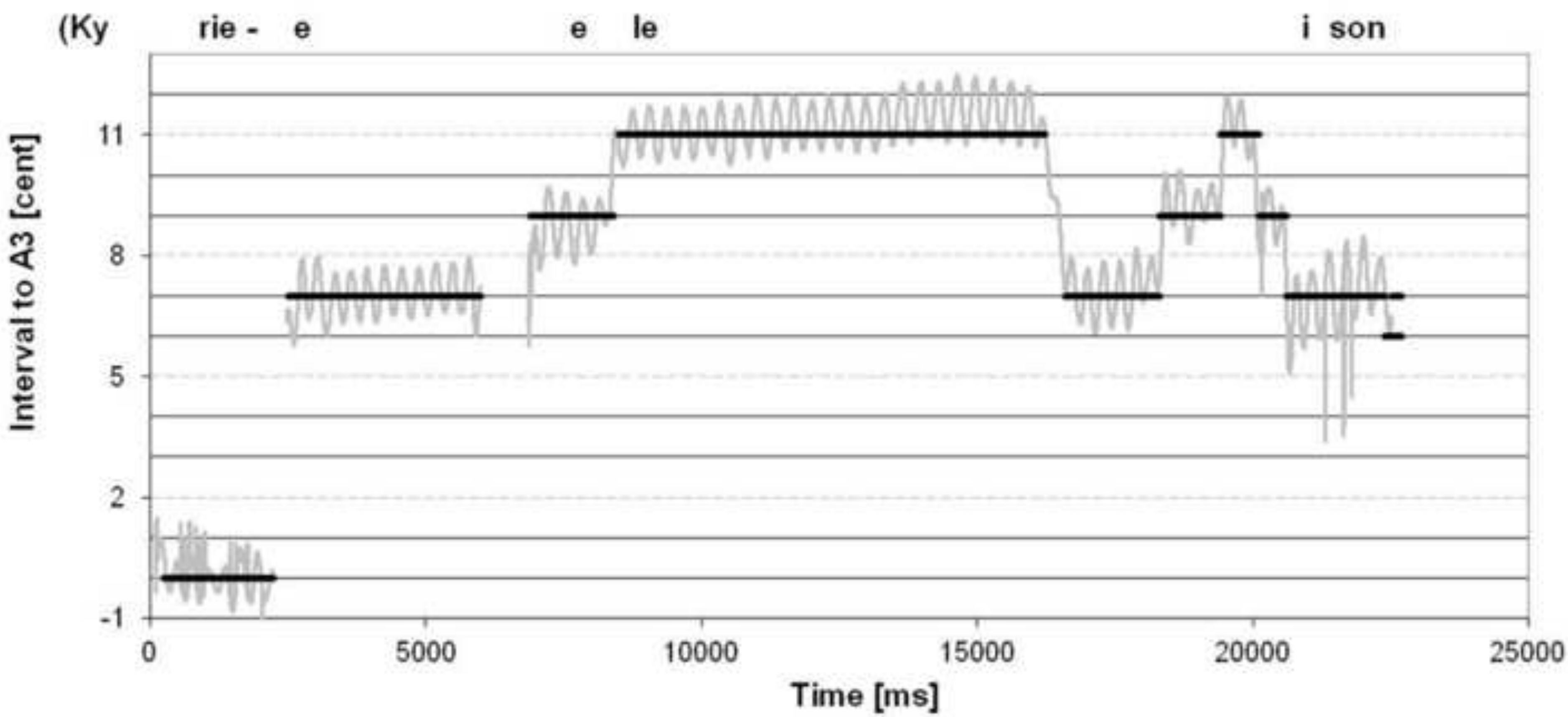


Figure 1

Figure(s)
Click here to download high resolution image

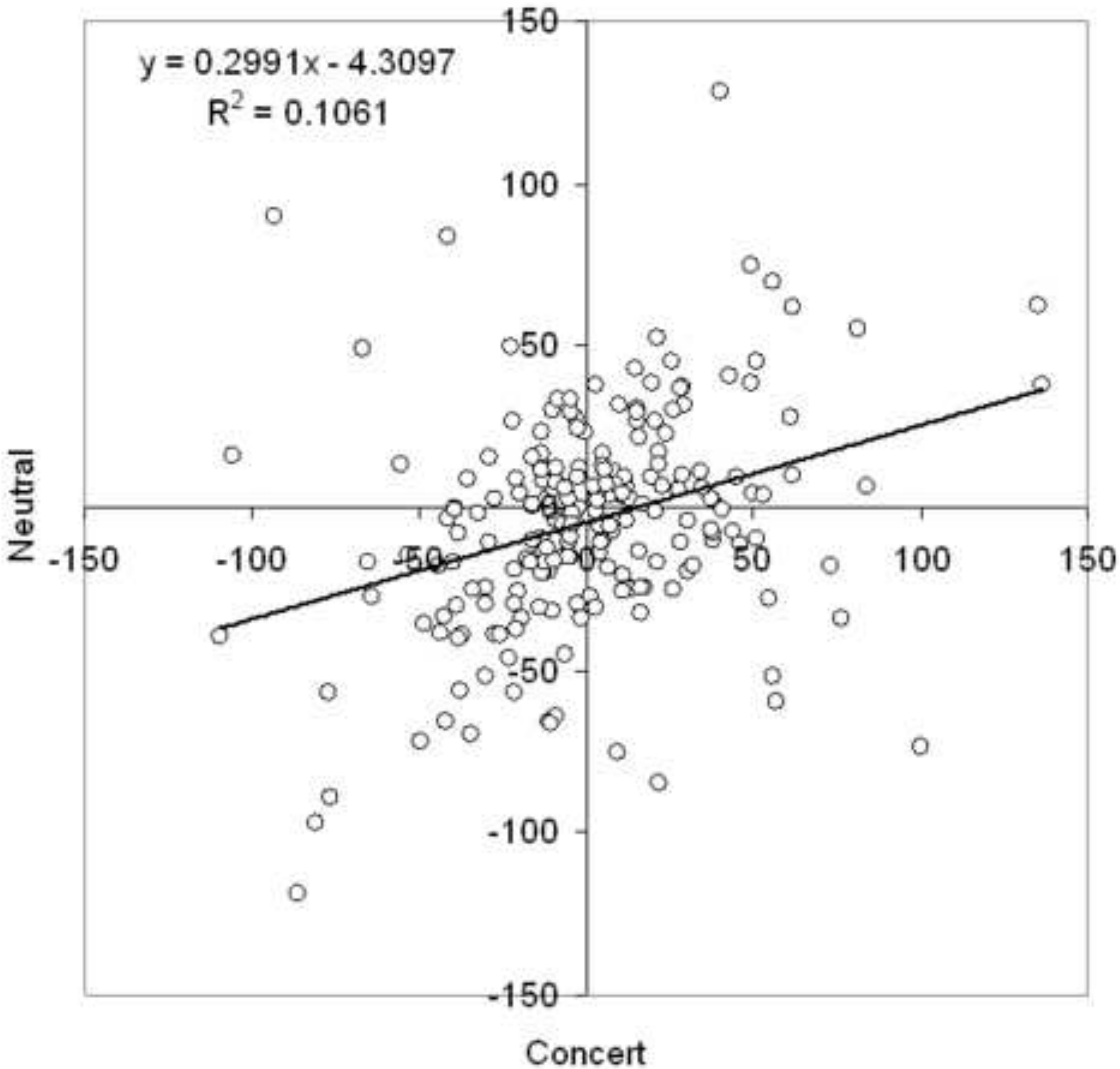


Figure 2

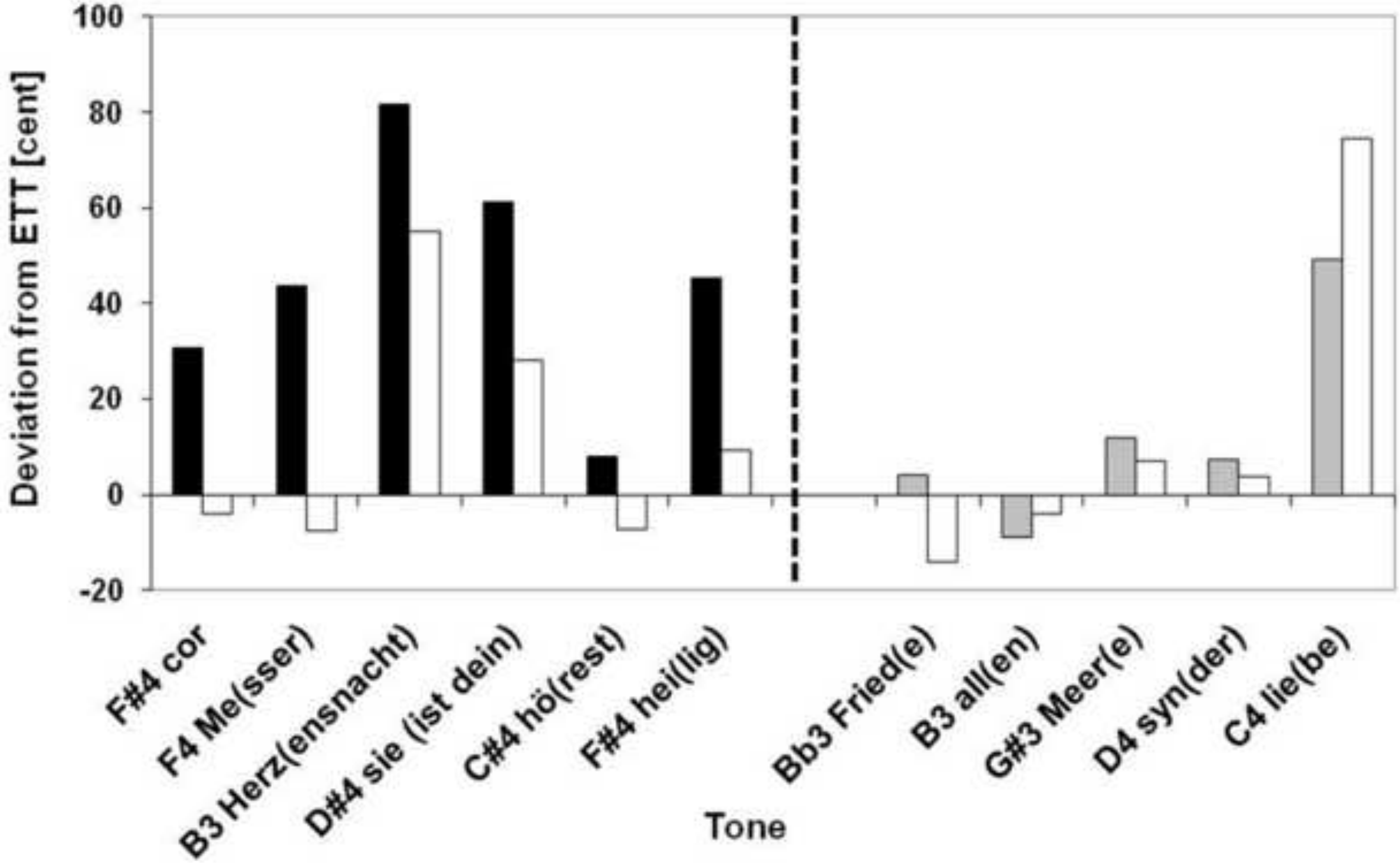

100
80
60
40
20
0
-20
Deviation from ETT [cent]
F#4 cor
F4 Me(sser)
B3 Herz(ensnacht)
D#4 sie (ist dein)
C#4 hö(rest)
F#4 hei(lig)
Bb3 Fried(e)
B3 all(en)
G#3 Meer(e)
D4 syn(der)
C4 lie(be)
Tone

Figure 3

Figure(s)
Click here to download high resolution image

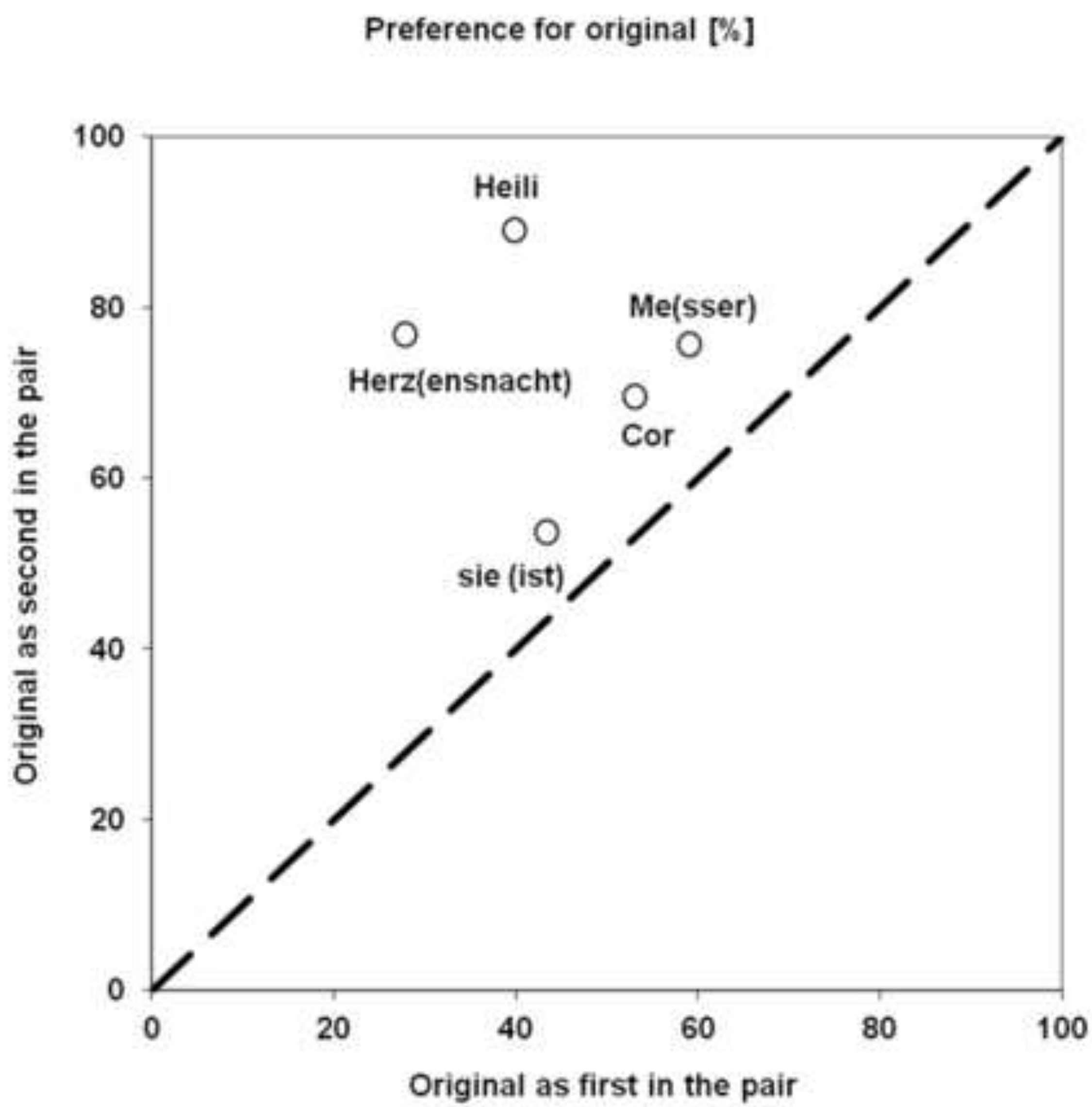


Figure 4