

Pitch Evaluations in Traditional Solo Singing: Comparison of Methods

Rytis Ambrazevičius¹, Robertas Budrys²

Faculty of the Humanities, Kaunas University of Technology, Lithuania

¹rytisam@delfi.lt, ²budrys@super.lt

ABSTRACT

Problems of pitch evaluations from pitch tracks obtained from computer aided acoustical analysis are considered; case of monophonic vocal performance is examined. The importance of limited jnd on the adequate desirable precision of the evaluation is noted.

Three methods of pitch evaluations were applied. First, pitches of one Lithuanian traditional vocal solo performance (six melostrophes) were independently evaluated manually from Praat-aided logf₀ tracks by three subjects. From these data on individual pitches, evaluations of musical scales averaged across the entire performance were also derived. Second, the evaluations of musical scales were repeated based on logf₀ histograms compiled from Praat readings. Third, software NoteView for automated pitch extraction and integral evaluation was applied. NoteView was chosen since it is considered one of the best programs for this purpose.

Evaluations of individual pitches by the three subjects (1st method) differed by 6.5 cents (here and hereafter averaged values are presented). However, for the degrees of musical scale, the difference dropped to 1.6–3.4 cents, depending on the range of sound durations (IOIs) considered. In comparison, the other two methods gave considerably inferior results (deviations from the semi-manual evaluations of the musical scale): 6.0–10.0 cents for histograms (2nd method) and 3.9–7.9 cents for NoteView (3rd method).

The semi-manual method of pitch evaluation, though time-consuming, is still more acceptable than the two automated methods considered; unless precision of 4.0–9.0 cents or worse is sufficient. The reasons (need for subjective decisions, e.g., on target pitch, etc.) are discussed.

I. INTRODUCTION

Reliability of pitch evaluations is important for different purposes including studies of pitch perception in general, as well as those of intonation and musical scales. The particular importance regarding ethnomusicological research should be noted. First, there is abundant evidence of systematic differences between the traditional music scales and twelve-tone equal temperament. It was noticed early on in the period considered to be the “birth of ethnomusicology” (Ellis, 1885; Stumpf, 1901; Gilman, 1908; Abraham & Hornbostel, 1909–1910; etc.). Second, different types of folk intonation are characteristic of different stages of mode development (e.g., Alexeyev, 1986). Third, certain dynamic aspects of musical scales (dependence of intonation on sound duration, melodic context, and other parameters and features) can be traced (Fyk, 1994; Friberg, Bresin, & Sundberg, 2006, p. 151; Ambrazevičius & Wiśniewska, 2008). Thus, specific techniques of the discussed evaluations should be developed.

II. PROBLEMS OF PITCH EVALUATIONS

When estimating pitches from graphs or other data of acoustical analysis, certain questions arise. First, how can we be confident of our pitch evaluations – are the results reliable? Second, should we strive for ultimate precision; what level of precision is adequate?

We will concentrate on these two questions, although surely there are more. For instance, the case of inharmonic sounds should be mentioned: is pitch evaluation from acoustical graphs possible at all in this case (Schneider, 2001, 2002)? Also, the very possibility of evaluation of pitch (subjective parameter) based on the examination of f₀ (objective parameter) might seem questionable. Well-known dependence of pitch on SPL or (maybe slightly less known) dependence of pitch on timbre (e.g., Vurma & Ross, 2007) could strengthen those doubts.

Concerning these issues, we should note that the sounds we examine (vocal performance) are actually harmonic and the ranges of change of timbre and loudness are too narrow to result in significant relative pitch deviations. The absolute pitch deviations from their logf₀-equivalents are not relevant in our case, because the musical scales are of most importance, i.e., the (adequately) precise intervals between the pitches, and not the precise pitches themselves.

A. Pitch jnd

The question of adequate precision is related to pitch jnd. Usually pitch jnds considerably exceed the values of several cents only found under the best listening conditions, limited ranges of durations, pitches, and loudness, certain musical contexts, and simple tone arrangements (successive tones, with no interference) characterized by the sharpest perception (Hess, 1983; Parncutt & Cohen, 1995, p. 863; Zwicker & Fastl, 1999, p. 185; etc.). For instance, shortening of sound duration raises jnd significantly (Zwicker & Fastl, 1999, p. 186); interfering interpolated tones can even result in the confusing of pitch classes (Deutsch, 1999, p. 391–396).

B. Problem of Unsteady f₀

Brief changes of logf₀ contribute to a certain fuzziness of the perceived pitch and make the evaluation of pitch problematic. It has been demonstrated, for instance, in experiments of perception of short glides (Nábělek, Nábělek, & Hirsh, 1970) or vibrato tones (Brown & Vaughn, 1993). It was concluded that the perceived pitch corresponds to the average frequency of the instances. However, it is unclear whether arithmetic or geometric means should be applied. Experiments by Rossi (1978) on prosody and falling glissando showed suitable pitch

¹ The study is supported by European Social Fund (Global Grant)

estimates corresponding to the frequency at 2/3 of the entire duration. Rützel & Ross (1985, p. 18) noted similarly on convex shapes of $\log f_0$ tracks: “We tried roughly to determine extreme frequencies of the convex and then to take the arithmetic mean” (which in many cases corresponds to the time point at two thirds of the duration of the note)”. However, “the reliability of this procedure cannot be taken for granted [...] and in the future a special experiment is needed in order to investigate this phenomenon” (ibid.). Researchers d’Alessandro & Castellengo (1994), and d’Alessandro & Mertens (1995) studied perception of short segments of vibrato tones and concluded the effect of memory decay on perceived pitch.

Therefore, sufficient precision of pitch evaluations for a study of pitch phenomena (musical scales and intonational aspects) in traditional vocal solo performance (characteristic of quite unstable fundamental frequency) could be considered to range from several cents to some 10, 20, or even more cents, depending on the specific case and the evaluation task. Nevertheless, the lower values of pitch jnd (i.e., 3–5 cents or so) could be fixed as a safe limit applicable to all or almost all cases.

III. METHODS

A Lithuanian harvest song *Vaikščiojo tėvulis* performed by renowned Lithuanian female folk singer Marė Navickienė has been chosen for the analysis (Fig. 1). The recording of the song was made in 1956 and contained 14 melostrophes featuring complicated semi-free rhythm and abundant ornamentations. The first six melostrophes were analyzed in three ways to compare different methods of pitch evaluations.

Tempo rubato ≈ 58

Figure 1. *Vaikščiojo tėvulis*. Transcription of the first melostrophe (Četkauskaitė, 2006, p. 28).

A. Semi-Manual Evaluations

Pitches and onsets of each sound of the song were measured. A small number of grace notes were not considered because of the crude uncertainty of pitch. Computer software for acoustical analysis PRAAT was applied. Perceived (integral) pitches of tones were estimated from continuous tracks of “objective pitch” (log frequency) automatically transcribed by the software. The following methods were applied (see also Ambrazevičius 2005–2006, p. 66–67).

1. Gliding onset and offset of a tone were omitted from consideration. Pitch of the remaining quasistationary segment was measured. In the case of vibrato, several full cycles were considered.

2. In the case of perceived slow intratonal pitch change (i.e., mostly characteristic of prolonged tones), a “target tone” was specified, and the corresponding segment of pitch track was considered.

3. In the case of considerable and irregular intratonal pitch change, pitch of short segments chosen from the track was aurally compared to the pitch of the entire tone. This technique is applicable, provided more or less stationary short segments are found in the track. Also, short segments of monotonically ascending or descending pitch track were accepted for the estimation. In this case the aspect of memory decay in perceptual pitch evaluation (e.g., d’Alessandro and Mertens, 1995) was taken into account.

4. The sound onsets were used for the evaluation of sound durations, or, to be precise, for the evaluation of IOIs (Inter-Onset-Intervals). It is well known that the fixing of the perceived sound onsets is relatively simple and reliable only in the case of short and prominent attacks. Generally it is quite problematic (see, for example, the studies on perceptual attack time – PAT: Vos and Rasch, 1981; Gordon, 1987, etc.). Nevertheless, this was considered to be not a relevant problem for the current study since very rough approximations of IOIs are sufficient in this case.

Three subjects (the authors and a colleague) measured pitches and onsets of all sounds in melostrophes 1–3 independently. After the results were collated, typical shortcomings were revealed and discussed. Then the procedure was repeated with melostrophes 4–6. Also, one subject repeated her measurements of melostrophe 3 after several days (see later in Fig. 3).

The occurrences of scale degrees in a melostrophe were averaged across the melostrophe to obtain the averaged musical scale in the melostrophe. This procedure was repeated with all six melostrophes, for all three subjects. The data was then applied to compare the individual findings of the subjects as well as to compare different pitch evaluation methods (see Results).

B. Histograms

The histogram method is based on the notion that quasistationary pitch segments corresponding to pitch categories are significantly longer than transitions, glissandos, glides and nonstructural sounds. This method is most suitable for the estimation of musical scale of the entire performance. Of course, it could be applied to short extracts or even single sounds as well, however, this technique would be even more time-consuming than the semi-manual evaluations just described.

A histogram accumulating $\log f_0$ track readings is designed (this can be automated and takes only a short time). Discriminate peaks of the histogram show the scale degrees (pitch categories) and valleys show boundaries between the categories (Fig. 2; top).

The advantages of this method are: (1) significant time savings, (2) simplicity, and (3) objectivity of the results which are based only on the statistics of $\log f_0$ track and not on the subjective techniques of evaluation. The disadvantage of the method is insufficient precision of results since the histogram bars show quite wide pitch intervals (bin-widths). In our case,

the intervals were 8–12 cents; attempts to narrow them led to “noisiness” of the histogram which masked the peaks.

Several procedures could be applied to overcome this disadvantage – to smooth or spline the “noisiness”. For instance, the $\log f_0$ track can be analyzed several times, with a small gradual shift of the bins (say, by 1 cent). The results for a certain pitch category can be obtained from the histogram which has the highest peak for that category. This principle of a sliding window can be also applied cumulatively, i.e., composing all the readings of the shifted histograms onto a single graph. The resulting curve (Fig. 2; bottom) shows many more details than the ordinary histogram. The mode or median for the certain pitch category could be also roughly evaluated applying the corresponding statistical methods for grouped data.

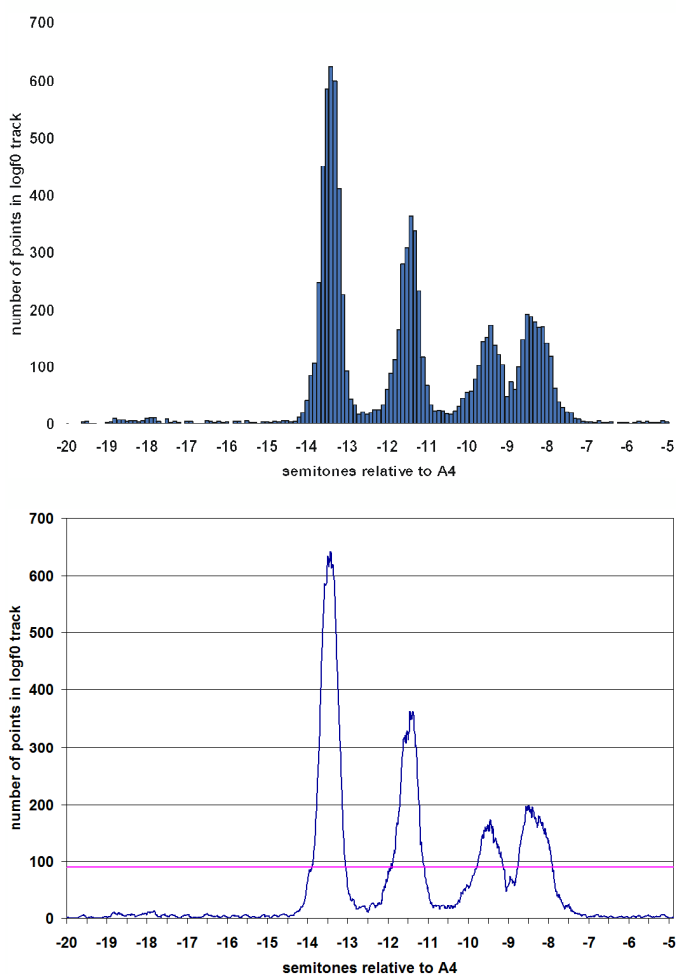


Figure 2. Pitch histograms; 4–6 melostrophes. Top: bin-width 10 cents. Bottom: sliding window of 10 cents is applied. Cutoff at 90 points displayed.

The histogram method can be applied in different ways.

1. The median, mean or maximum (mode) of a peak can be chosen alternatively as a resultant value of a scale degree. In the cases of median and mean, separation of different scale degrees is needed. This can be done, for example, by cutting the distribution curve at some level (Fig. 2, bottom) and considering only the instances above it.

2. Some parameters can be adjusted, such as time resolution of $\log f_0$ track, window size, and the level of the cutoff

(obviously, the cutoff does not affect calculations of the maximum).

3. To obtain the musical scale, (1) entire $\log f_0$ track can be analyzed or (2) its separate segments (e.g. single melostrophes) can be analyzed and then the results for each segment can be averaged. The latter technique could help to avoid “blurring” of peaks in the case of gradual transposition of the scale, for instance, gradual raise of the scale characteristic of unaccompanied folk singing.

The histogram method was applied to melostrophes 4–6. The parameters were varied. The $\log f_0$ track information was extracted using PRAAT software; several different time resolutions were applied, from .005 to .05 s, to determine whether it has any influence on results.

C. NoteView

NoteView is a software tool that can “take as its input a sound recording of a single line instrument <...>, parse the notes of the performance into a list of events (that could be inspected in both tabular and graphic forms), and to provide a comparison of this event list with an event list of another performance (also reported via tables and graphs)” (Gunawan & Schubert, 2010a, p. 25). The parsing is based on $\log f_0$ track deviations and pitch strength (Gunawan & Schubert, 2010a, p. 26). SWIPE’ algorithm (Camacho, 2007) is used to estimate $\log f_0$ information and pitch strength (Gunawan & Schubert, 2010a). We used the parsing function of NoteView, so the semi-manual method was partially automated.

Every single event corresponds to one perceived sound, and a set of parameters is estimated for each event. Only the parameters related to temporal position of event (*onAttack*, *on* and *off*) and to its pitch (median and mean of within-event fundamental frequencies, expressed in semitones and cents, i.e., the parameters *MIDI* and *mean*) were considered (Table 1; for details, see Gunawan & Schubert, 2010b). Scale degrees were assigned to each event manually.

Table 1. Demonstration of NoteView: first ten events of *VaikščiŃojo tėvulis*.

Event #	onAttack	on	off	MIDI	mean
1	0.91	0.97	1.24	49.626	49.707
2	1.32	1.69	2.47	55.296	55.333
3	2.73	2.82	3.11	55.091	55.176
4	3.38	3.38	4.21	60.726	60.739
5	4.27	4.28	4.43	57.371	57.239
6	4.45	4.45	4.94	55.146	55.19
7	5.04	5.04	5.43	58.931	58.905
8	5.44	5.44	5.6	57.846	57.858
9	5.72	5.72	5.91	60.761	60.75
10	5.95	5.95	6.28	59.211	59.269
...

This method was applied to melostrophes 4–6. Just as in the semi-manual method, the occurrences of certain scale degrees (the values of pitch median or pitch mean) were averaged across melostrophes 4–6 to obtain the averaged musical scale.

IV. RESULTS

A. Semi-Manual Evaluations

After the revelation of typical shortcomings in the measurements of melostrophes 1–3, the measurements of the succeeding melostrophes 4–6 were more precise (Fig. 3). Therefore, only the results obtained from melostrophes 4–6 will be considered.

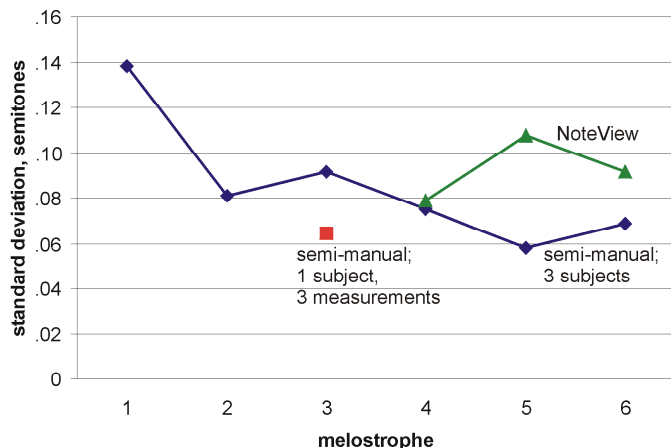


Figure 3. Averaged standard deviations between the evaluations of individual pitches given by three subjects (semi-manual experiment). Also, the averaged (absolute) deviations between the NoteView readings and semi-manual evaluations are shown; for melostrophes 4–6.

Fig. 3 shows that the evaluations of individual pitches by the three subjects (method of semi-manual evaluations) differed by roughly 6–8 cents, on average, for separate melostrophes. Averaged pooled value for melostrophes 4–6 is 6.5 cents. It is important that the deviations between the evaluations by the three subjects diminish in the case of longer durations of pitches (Fig. 4). This is in accordance with the discussed dependence of pitch jnd on duration. The standard deviations of the pitch evaluations were less than 10 cents starting from app. 450 ms. Consequently, decisions regarding individual pitches can be considered acceptably precise only for quite long sounds (if making no additional time consuming attempts and applying no additional intricate methods of evaluation).

However the decisions regarding musical scales averaged across the entire performance can be considered valid when applying the data of significantly shorter sounds. The deviation of pitch (i.e., scale degree) evaluations were shown to be noticeably less than 10 cents even when all short ornamental sounds were included in the analysis; the average of the deviations dropped to 3.0 cents. These deviations were found to be different for different scale degrees (see “all notes” in Fig. 5). Interestingly, the 1st and 4th scale degrees showed the least deviations. This probably means that their intonations were the most stable and easily measured. This is in accordance with the tonal structure of this particular song. As evident from Fig. 1, the structure anchors on the two scale degrees (1st and 4th) which are traditionally considered by ethnomusicologists as the lower and upper tonics, forming the so-called quart-tonic structure.

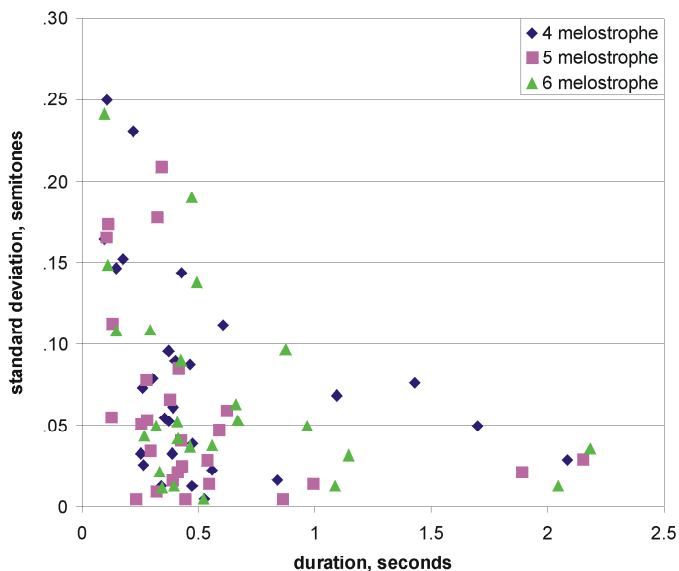


Figure 4. Standard deviations between the evaluations of individual pitches given by three subjects (semi-manual experiment): dependence on sound durations (IOIs).

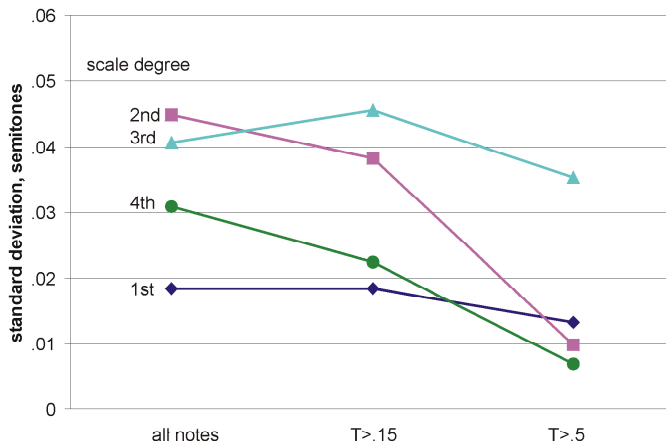


Figure 5. Averaged standard deviations between the evaluations of musical scales given by three subjects (semi-manual experiment): dependence on sound durations (IOIs) and scale degrees. Results for melostrophes 4–6 are shown.

The deviations of pitch (scale degree) evaluations among the three subjects were typically less when only longer pitches were considered (Fig. 5). For instance, when only the pitches longer than 500 ms were taken into account, all scale degrees except the third showed deviations of approximately 1 cent only. Consequently, if this extreme level of precision (and reliability) is needed, only the prolonged pitches should be estimated. However, a sufficient number of pitches should be considered. For instance, the example of the 3rd scale degree in Fig. 5 shows that the very small number of its occurrences in melostrophes 4–6 (two pitches longer than 500 ms) results in significantly rougher estimations. Another issue is musical context: if we suspect the scale degrees to be intoned differently in different musical contexts, we should compose the appropriate sets of pitch data.

B. Comparison of Methods: Histograms and Semi-Manual Evaluations

The following settings were used to compose the data applicable for the consecutive tasks: time resolution .05 s, sliding window 50 cents, and cutoff at 47 points (when applied to melostrophes 4–6). These settings showed relatively small deviations from the findings of the semi-manual experiment. It should be noted that analysis of the results obtained from histograms showed that choice of settings of different parameters did not considerably affect the results. Additionally, the results show no clear tendencies of dependence of the deviations on the settings. Thus, the settings resulting in (seemingly) the least deviations were used.

The averages of histogram peaks corresponding presumably to the averaged values of certain scale degrees were compared to the averages of the semi-manual evaluations (Fig. 6). Additionally, the corresponding averaged deviations between the evaluations of the subjects in the semi-manual experiment (i.e., averages of the data in Fig. 5) are presented. The conclusion is that the histogram method gives pitch estimations of scale degrees significantly (2–3 times) worse than the semi-manual method, even with suitable settings applied.

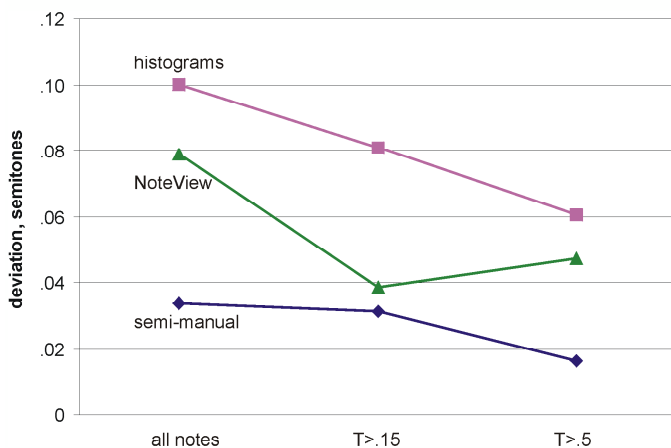


Figure 6. Absolute deviations between the evaluations of musical scales; results averaged across the four scale degrees and for melostrophes 4–6.

Choice of different settings resulted in the deviations from the semi-manual evaluations, mostly from 6 to 15 cents, but up to 34 cents in certain cases.

C. Comparison of Methods: NoteView and Semi-Manual Evaluations

Analysis of NoteView results showed that this software missed a total of 19 notes in melostrophes 4–6, i.e., roughly 21 % of events. Collation of the pitches of mutual notes (occurring both in the semi-manual evaluations and in the NoteView readings) shows the differences of approximately 9.3 cents on the average; roughly the same for median and mean (Fig. 7).

If the non-mutual notes are not omitted from the analysis, and pitches of the scale degrees are estimated, the difference drops slightly to almost 8 cents (Fig. 6). Consideration of longer sound durations gives better results; still they are worse than the results of the semi-manual experiment.

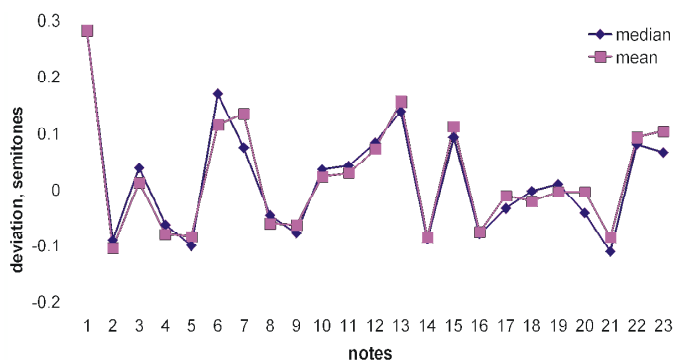


Figure 7. Melostrophe 4: deviations between the NoteView readings and semi-manual evaluations.

V. DISCUSSION

The idea of automated pitch evaluation is attractive because of significant saving of time resources. If imprecision of some 10 cents does not matter, the considered automated methods can be applied. Nevertheless, the semi-manual method of pitch evaluation, though time-consuming, is preferable if more precise evaluation comparable to the lower values of jnd is required.

The decision on the requirements depends on the phenomenon studied. For example, if we are to compare the instances differing by 30 or 40 cents for the purpose of objectivizing and concluding the rough difference, the automated methods discussed probably would be sufficient. If we are interested in differences of some 10 cents, then most probably we should choose the semi-manual evaluations.

This holds for the examined type of performance, i.e., characteristic of significant unsteadiness of logf0 track. For relatively steady pitches, more precise evaluations could be carried out applying the same methods.

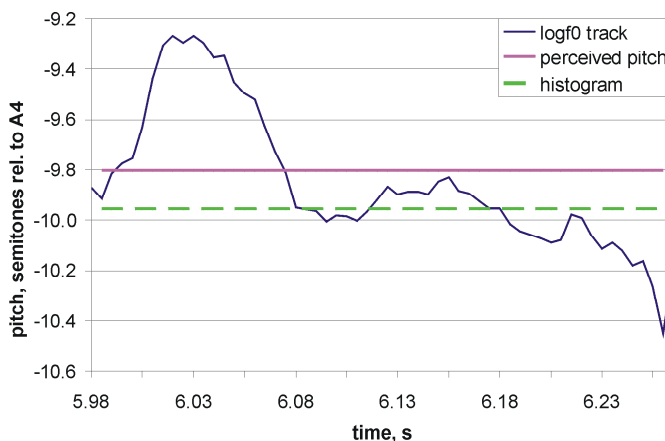


Figure 8. Pitch track of one note (pitch event). Perceived pitch and evaluation given by histogram are depicted.

There are several reasons why the deviations discussed are relatively large for the automated methods. We find a couple of them to be the most important. First, when measuring semi-manually, we make subjective decisions on the “target” portions of pitch events. For instance, starting glides, or certain weaker segments are considered as supplements to the main body of a pitch and they are omitted from the analysis. In contrast, the automated methods include them into the analysis,

so the resultant pitch evaluation is generally shifted (Fig. 8). Consequently and somewhat paradoxically, the seeming objectivity of a method may be its drawback.

The automated methods also include “technical garbage” in the analysis, i.e., accidental software misinterpretations (due to problems of settings, noise, etc.).

The deviations of NoteView results could be also attributed to some shortcomings of the SWIPE’ pitch extraction algorithm implemented into the NoteView software.

REFERENCES

- Abraham, O., & von Hornbostel, E. M. (1909–1910). Vorschläge für die Transkription exotischer Melodien. *Sammelbände der Internationalen Musikgesellschaft*, 11, 1–25.
- d’Alessandro, C., & Castellengo, M. (1994). The pitch of short-duration vibrato tones. *Journal of the Acoustical Society of America*, 95(3), 1617–1630.
- d’Alessandro, C., & Mertens, P. (1995). Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language*, 9, 257–288.
- Alexeyev, E. (1986). *Rannefol’klornoe intonirovanie. Zvukovysotnyj aspekt*. Moscow: Sovetskij kompozitor.
- Ambrazevičius, R. (2005–2006). Modelling of scales in traditional solo singing. *Musicae Scientiae*. Special Issue "Interdisciplinary Musicology", 65–87.
- Ambrazevičius, R., & Wiśniewska, I. (2008). Chromaticisms or performance rules? Evidence from traditional singing. *Journal of Interdisciplinary Music Studies*, 2(1&2), 19–31.
- Brown, J. C., & Vaughn, K. (1993). Pitch center of musical sounds with vibrato. *Journal of the Acoustical Society of America*, 94(3), 1860–1860.
- Camacho, A. (2007). *SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech and Music* [doctoral dissertation]. Gainesville, FL: University of Florida.
- Četkauskaitė, G. (Ed.) (2006). *Dzūkų dainininkė Marė Kuodžiūtė-Navickienė*. Vilnius: Lietuvos muzikos ir teatro akademijos muzikologijos instituto etnomuzikologijos skyrius.
- Deutsch, D. (1999). The processing of pitch combinations. In D. Deutsch (Ed.), *Psychology of Music* (2nd edition) (pp. 349–411). San Diego, London: Academic Press.
- Ellis, A. (1885). On the musical scales of various nations. *Journal of the Royal Society of Arts*, 33, 485–527.
- Friberg, A., Bresin, R., & Sundberg, J. (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology*, 2(2–3), 145–161.
- Fyk, J. (1994). Static and dynamic model of musical intonation. In A. Friberg et al. (Eds.), *SMAC 93. Proceedings of the Stockholm Music Acoustics Conference. July 28–August 1, 1993* (pp. 89–95). Stockholm: Royal Swedish Academy of Music.
- Gilman, B. I. (1908). Hopi songs. *A Journal of American Ethnology and Archeology*, 5.
- Gordon, J. W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America*, 82(1), 88–105.
- Gunawan, D., & Schubert, E. (2010a). NoteView: A computer program for the analysis of single-line musical performances. *Acoustics Australia*, 38(1), 25–30.
- Gunawan, D., & Schubert, E. (2010b). *NoteView (Version 0.5)*. Sydney: The University of New South Wales.
- Hess, W. (1983). *Pitch determination of speech signals. Algorithms and devices*. Berlin: Springer Verlag.
- Nábělek, I. V., Nábělek, A. K., & Hirsh, I. (1970). Pitch of tone bursts of changing frequency. *Journal of the Acoustical Society of America*, 48(2), 536–553.
- Parncutt, R., & Cohen, A. J. (1995). Identification of microtonal melodies: Effects of scale-step size, serial order, and training. *Perception and Psychophysics*, 57(6), 835–846.
- Rossi, M. (1978). La perception des glissando descendants dans les contours prosodiques. *Phonetica*, 35, 11–40.
- Rüütel, I., & Ross, J. (1985). *A study of pitch contours and the scale structure in Votic folk music. Preprint KKI-37*. Tallinn: Academy of Sciences of the Estonian SSR, Division of Social Sciences.
- Schneider, A. (2001). Sound, pitch, and scale: From “tone measurements” to sonological analysis in ethnomusicology. *Ethnomusicology*, 45(3): 489–519.
- _____ (2002). On tonometrical and sonological analyses of “exotic” instruments: From Stumpf’s measurements to the present. In G. Berlin & A. Simon (Eds.), *Music archiving in the world. Papers presented at the Conference on the occasion of the 100th anniversary of the Berlin Phonogramm-Archiv* (pp. 247–257). Berlin: Verlag für Wissenschaft und Bildung Staatliche Museen Berlin.
- Stumpf, C. (1901). Tonsystem und Musik der Siamesen. *Beiträge zur Akustik und Musikwissenschaft*, 3, 69–138.
- Vos, J., & Rasch, R. (1981). The perceptual onset of musical tones. *Perception & Psychophysics*, 29(4), 323–335.
- Vurma, A., & Ross, J. (2007). Timbre-induced pitch deviations of musical sounds. *Journal of Interdisciplinary Music Studies*, 1(1), 33–50.
- Zwicker, E., & Fastl, H. (1999). *Psychoacoustics. Facts and models*. Berlin, Heidelberg: Springer-Verlag.