Barış Bozkurt
Bahçeşehir University
Electrical and Electronics Engineering
Department
Çırağan Caddesi Osmanpaşa Mektebi
Sokak No: 4–6, 34353
Beşiktaş, İstanbul, Turkey
baris.bozkurt@bahcesehir.edu.tr

A System for Tuning Instruments Using Recorded Music Instead of Theory-Based Frequency Presets

Abstract: Musical instrument tuners are devices that help musicians to adjust their instruments such that the played notes have the desired fundamental frequencies. In a conventional tuner, the reference tuning frequencies are preset, where the presets are obtained from tuning (musical scale) theory, such as twelve-tone equal temperament, or are user-specified temperaments. For many kinds of music in oral traditions, especially nonwestern music, widely accepted theoretical presets for tuning frequencies are not available because of the use of non-standard tunings. For such contexts, the "reference" is a master musician or a recording of a master musician. In this article, a tuning method and technology are presented that help the musician to tune the instrument according to a given (user-provided) recording. The method makes use of simultaneous audio and visual feedback during the tuning process, in which novel approaches are used for both modalities. For audio feedback, loopable stable frames, obtained automatically from the recording, are looped and played continuously. For visual feedback, a superimposed plot of the auto-difference functions is displayed instead of the conventional tuner's approach of detecting frequencies and displaying the amount of frequency difference between the input and the reference.

A musical instrument tuner is a well-known device that is used when adjusting a musical instrument so that the sounds are produced at the desired vibration frequencies (f_0) when the instrument is played. Tuners exist in many forms: as compact hand-held electronic devices, or as software running on personal computers, effect processors, or even mobile phones. Due to their economic value, the literature on musical tuners is largely composed of patents on hand-held devices (i.e., circuit designs).

Conventional tuners help the process of adjustment (usually a mechanical adjustment, such as changing the tension of a string or the length of an acoustic tube) by displaying the amount of deviation between a played pitch and a desired pitch. Then the musician, using the visual display, tries to minimize the deviation. In the design of conventional tuners, especially software-based types, the main signal processing problem is the f_0 estimation. Once f_0 is estimated, the remaining problem is comparing f_0 to a preset value and displaying the difference. f_0 estimation is one of the most frequently studied topics in audio signal processing, with numerous methods and plenty of actual use in speech and audio processing applications.

Computer Music Journal, 36:3, pp. 43–56, Fall 2012

As presented herein, our approach deviates highly from conventional tuners (both hardware and software) and does not propose a new f_0 estimation method; on the contrary, it proposes a new approach that avoids the f_0 estimation during the conventional tuning process. Our tuner is recording-driven, i.e., the reference is a user-provided recording. Because of this property, the design (which includes analysis of a recording and further comparison to frames of the recording) has almost nothing in common with hardware tuner systems. Therefore, neither a review of conventional tuners (e.g., patents on circuit designs), nor a review of f_0 estimation methods, will be presented here. For f_0 estimation, see Hess (1983), de Cheveigné and Kawahara (2002), and Camacho (2007).

Despite the fact that plenty of tuning tools and products are available, the need for tuners for traditional music is far from being met. For example, one particular case is the traditional makam music in Turkey (TMMT). In private communication of the author with masters of this music, the need for tuners was mentioned as an important deficiency in music technology that leads to difficulties in playing specific microtonal intervals, especially for beginner and intermediate-level performers. In this context, it is important that the tuner helps tune all possible tones of the music style (as chromatic tuners do), not just a subset of notes corresponding to the open strings of a particular instrument.

The necessity of a tuning device is questionable for a master musician with a tuning fork. A learned musician may be expected to be able to tune his or her instrument correctly without the need of a tuner. In various traditional music cultures, learning music also involves learning to tune the instrument. Tuning devices are still useful in many settings, however, even for master musicians, and especially when a quick tuning operation is needed, or when more than one fretted instrument will be playing together. For beginners, the tuning device facilitates the task of getting ready to start playing. In the case of traditional music with many microtonal intervals, the need for such help is great (except for fretless instruments, where tuning a few open strings is sufficient). A good example is the tuning of the kanun, a zither-like instrument used in TMMT. The kanun, having 24 to 26 triplets of strings (a total of 72 to 78 strings), is very difficult to tune. Without a tuning device, a master would need a considerable amount of time to tune the instrument, playing many different modes to check the tuning, and then iteratively correcting errors. In a musical setting with one kanun and a fretted instrument, such as a tanbur, the process becomes more complicated, because there may be tuning differences which are fine in solo playing but are problematic in a multi-instrumental setting. The tanbur, a long-neck lute, is one of the main instruments of TMMT. An important aspect of this instrument is that the frets are movable; the musician often needs to change the positions of several frets on a new instrument.

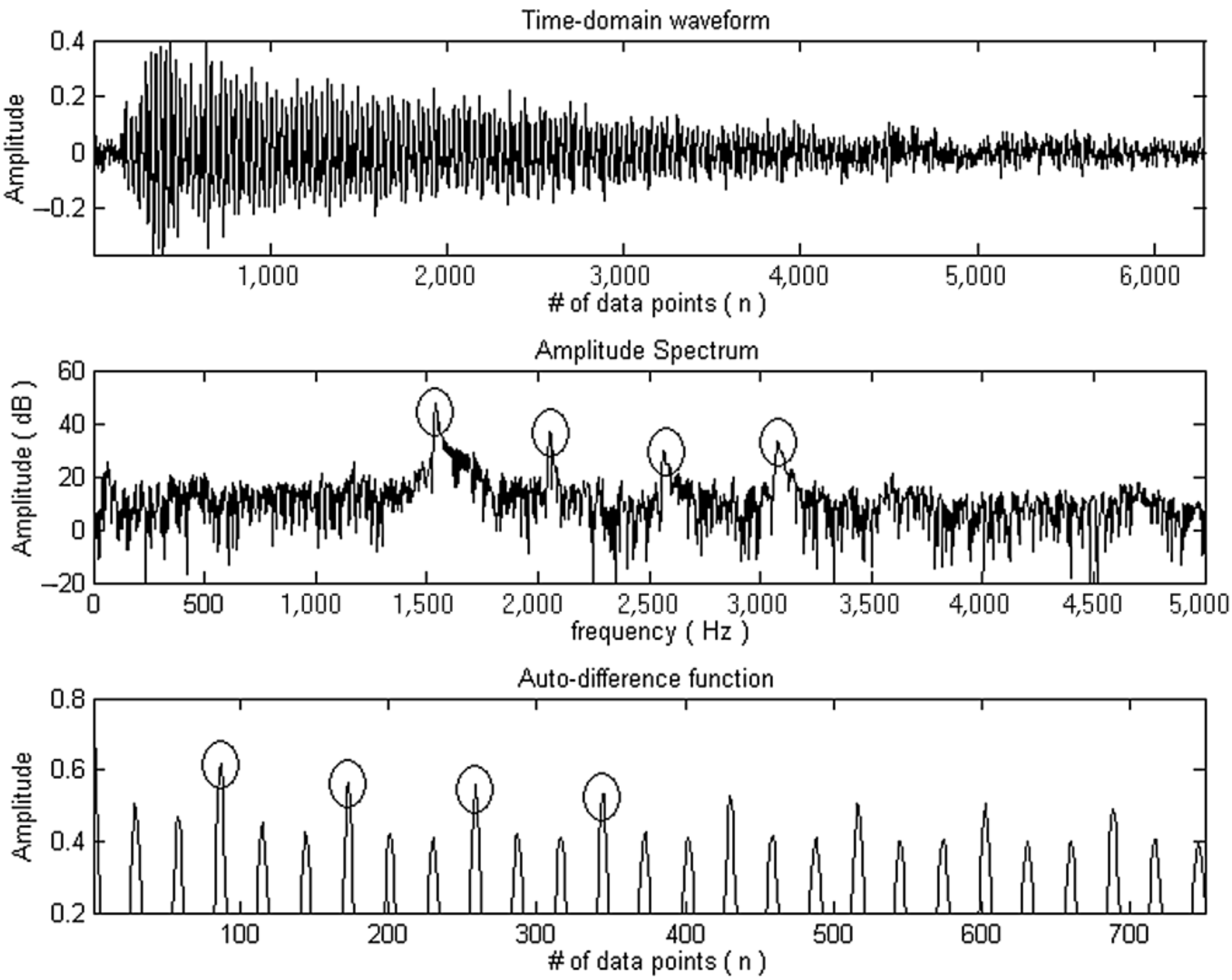
Today, many musicians of TMMT tend to use western-music tuners in various ways to overcome these problems, yet this risks modifying the tuning system due to technical deficiencies. Some master musicians claim that such modifications have already happened and that the tuning system of TMMT has been altered by the use of western-music tuners.

In many traditional music settings, theoretical presets for tuning frequencies are not available, because of the use of non-standard tuning systems. Tuning may vary regionally, according to this or that musician, or according to the instrument maker. For musical instruments to be played in such a context, the desired pitch information is not available through a tuning theory. The main source of information for desired pitches is a recording of a master musician. What is needed is a tuner system that accepts a reference recording as an input, to help the musician tune the instrument as in the recording. Our novel method is designed to serve this need, and provides a means of tuning without predefined, fixed tuning frequencies; it thus deviates highly from existing methodologies that tune to theoretically predefined or user-defined intervals or frequencies.

One might argue that instead the solution is to conduct research to define the tuning system, as in Tidhar, Mauch, and Dixon (2010), and that conventional tuners can be used with new presets. Such an approach would not be suitable for cultural contexts in which the standardization of tuning is intentionally avoided, however, because regional or individual variations are considered "colors" that should not be erased by theory. In the case of TMMT, and because of the previously mentioned reason, defining standardized tuning has been a controversial issue. Proposals vary from 17 to 79 tones in an octave (Yarman 2008). Although 24 tones are officially used in teaching and notation (Arel 1968), this does not conform to practice (Bozkurt et al. 2009). A commonly accepted tuning theory that conforms to practice is not available.

For tuning estimation based on histogram peaks, as shown in the Tarsos system (Six and Cornelis 2011), difficulties arise when all notes played do not show up as clear peaks in the pitch histogram. In addition, the influence of melodic direction may cause the pitch histogram peaks to become vague and introduce ambiguity in peak detection. It appears that a method going beyond these limitations and taking advantage of user interaction is useful. In our approach, we use automatic detection of stable frames in the signal, then use an automatic or manual method (depending on which of these two options the user chooses) to decide which frames are accepted or rejected. These frames are then used as reference signals in the tuning operation. Such an

Figure 1. A problematic example of fundamental frequency estimation: a tanbur recording at 44,100 Hz sampling frequency. Top: Time-domain waveform. Middle: Amplitude spectrum (four of the harmonics are marked with circles). Bottom: Auto-difference function (its computation is explained in detail in the section "The Application Part"). The period of the signal can be estimated from peaks marked with circles.



approach gives the freedom for individual variations in tuning, which is critically important in music styles such as TMMT, where masters from different schools may not prefer the same tuning references.

An additional aspect of our new design is the avoidance of the frequency estimation step; rather, we compare two sound signals in terms of frequency without measuring the frequency. Most frequency estimation methods, such as de Cheveigné and Kawahara (2002), first compute a new signal, such as the auto-correlation function or the magnitude spectrum, and then estimate the period of this new signal by various methods, such as peak picking (Hess 1983). One of the deficiencies of such systems is octave errors, especially when the fundamental component of the signal is relatively low in energy compared to other partials.

To support this idea, we present an example (see Figure 1) where the well-known and commonly used YIN algorithm (de Cheveigné and Kawahara 2002) fails to detect correctly the fundamental frequency. A visual observation on the auto-difference function provides a more reliable estimate. The fundamental frequency of a tanbur recording, with a 44,100 Hz sampling frequency, can be visually observed in Figure 1 to have a fundamental frequency around

507 Hz, given that the first peak of the auto-difference function is at index 87 and the frequency difference between the 3rd, 4th, and 5th harmonic peaks is about 500 Hz. The YIN estimation results, obtained from several signal windows from this data, vary within the range of 185.6 Hz to 187.6 Hz. This estimation is clearly wrong, and is probably due to the lack of strength in the first two harmonics, which is typical for some tanbur recordings. This type of error happens rarely and should not be overemphasized. A new method, however, that is observed to have better performance in such cases and is easier to implement, is proposed herein.

Here we propose that the intermediate signals used for f_0 estimation—for example, the auto-difference function (de Cheveigné and Kawahara 2002) of the reference and the input signals—can be provided as visual feedback to the user instead of frequency deviation information. As we show in our demonstrations, this visual feedback is sufficient to observe the discrepancy between the frequency of the input and that of the target signal. (Sound files and a movie of tuning are available online at akademik.bahcesehir.edu.tr/~bbozkurt/tuner.) Note that Figure 1 demonstrates the f_0 estimation problem. During the tuning operation, the user is not asked to perform visual peak picking to estimate f_0, but rather to check if several peaks of two signals match at the same time.

By avoiding the estimation step, the implementation is highly simplified, which is the main advantage of this approach. The system makes use only of rough frequency estimation via autocorrelation in detection of stability and filtering of the reference signals, which can be processed offline. Here, the accuracy of f_0 estimation is not as critical as in the actual tuning operation of a conventional tuner.

Without tuners, musicians perform the tuning operation by audio feedback: The reference and target sounds are played simultaneously and the musician adjusts the instrument such that the beating, resulting from the frequency difference of the two signals, disappears. Most tuners, interestingly, do not provide a sound output to imitate such a process. We think that audio output would be very beneficial for off-stage tuning, where sound output will not disturb a performance. For our tuner, the audio output is either the original reference signal frames, automatically detected, or a "loopable version" of these frames that will be continuous when looped. A novel method to obtain the loopable version is presented in the following sections.

In short, the novel aspects of our approach can be listed as (1) architecture of a tuning system that accepts a recording as input and helps the user to tune according to the recording, (2) an approach to tuning without a predetermined fixed frequency set, (3) a visual feedback mechanism to help frequency matching of two signals without frequency estimation, and (4) a novel method for obtaining loopable versions of stable audio segments to be used for aural feedback.

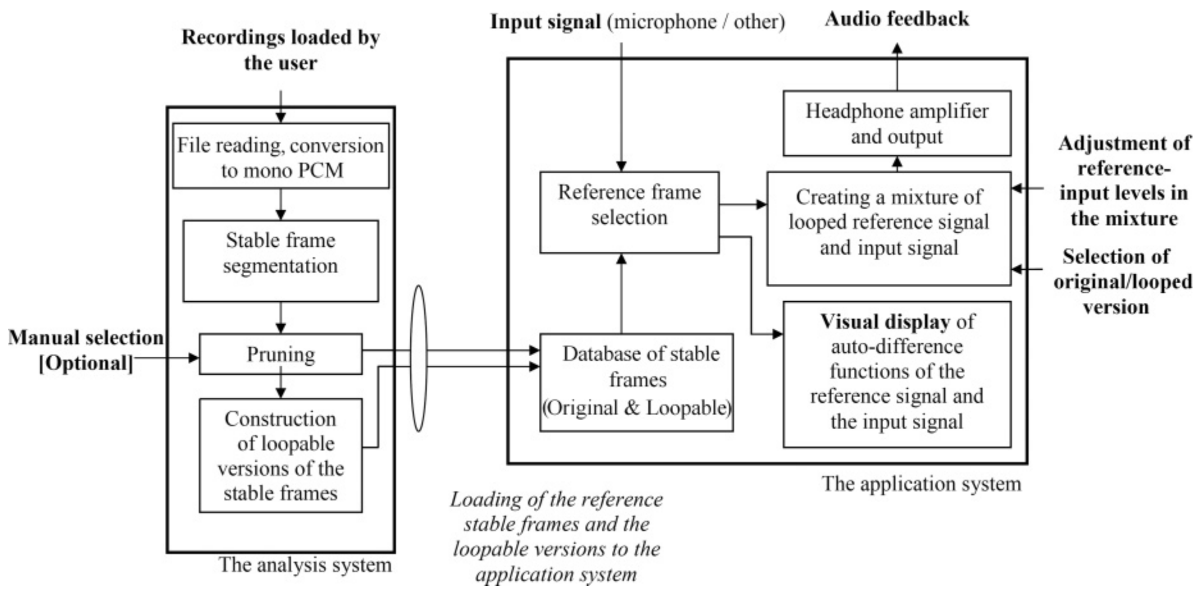
The Overall System Architecture

As shown in Figure 2, our tuner consists of an analysis system and an application system. The analysis part takes a reference recording (provided by the user in a format such as MP3 or WAV) and, in a fully automatic manner, extracts a list of "stable" frames that will serve as reference signals. A stable frame is a region of the waveform where vibration frequency is roughly constant. The recording should be monophonic, not polyphonic. Then, the set of reference frames are subjected to automatic or manual pruning, depending on the user's preference. As the final step of the analysis, loopable versions of the stable frames are obtained to provide continuous playback during the tuning operation.

The application part is designed to enable tuning using the stable frames extracted in the analysis. This part provides audio and visual feedback to the user about the distance of the input signal's frequency from the desired reference signal's frequency.

The complete system can be implemented to run on a full-size computer or on a hand-held device. Alternatively, it is possible to separate the two parts and let only the application part run on a hand-held device and the analysis part run on a full-size computer.

Figure 2. The overall system architecture.



The Analysis Part

The offline analysis module takes as input a user-provided sound file in a standard audio format; the recorded source should be monophonic. The sound data is read from the file and, if the given file contains stereo data, it is converted to mono by a simple amplitude normalization and addition operation. This one-dimensional sound data is then analyzed to extract stationary frames.

Stable-Frame Segmentation

Automatic stable-frame segmentation can be performed in various ways. One way to approach the problem is to treat it as a counter-problem to the problem of onset detection, where sudden changes in a signal are tracked. In onset detection, it is common practice to derive the onset strength signal (OSS; Holzapfel et al. 2010) from the original signal using a parametric time-series such as the temporal variation of amplitude or energy (Schloss 1985; Moore, Glasberg, and Bear 1997), the spectral deviation (in amplitude, phase, complex components of spectrum [Bello et al. 2005], or group delay [Holzapfel et al. 2010]), the derivative of fundamental frequency, or the combination of some approaches (Zhou and Reiss 2007; Holzapfel et al. 2010). In a stable-frame segmentation problem, the task is to segment portions where the OSS stays almost constant (i.e., within a specified dynamic range). Here, all proposed OSSs have a certain potential where the derivative of f_0 appears to be more suited to our problem of obtaining reference signals for tuning. In our system, stable-frame extraction is achieved by first estimating f_0 using a simple autocorrelation peak-detection method (Hess 1983), then partitioning audio into segments where a maximum pitch deviation is allowed to be a fixed, user-defined value (with a default set to 25 cents). The segments that last less than a user-defined duration (with a default set to 250 msec) are discarded. The accuracy of f_0 detection is not too critical for this step, as it is only used for detecting stationary portions. Further, f_0 estimates are also used for setting the cut-off frequencies for filtering, where a rough estimate is still sufficient.

Given the set of stable frames extracted from the signal, the next step is to prune the set. This is necessary because, as one would expect, several instances of the same note are likely to be detected. In addition, some out-of-tune notes would still be selected as stable segments but should be discarded. Our preferred method is to perform

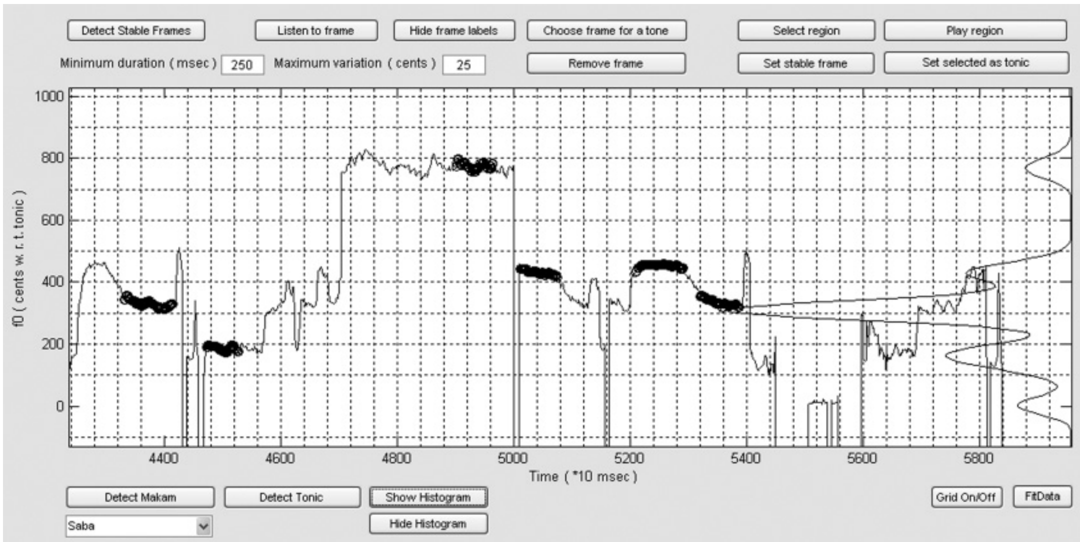


Figure 3. Snapshot of user interface for pruning, editing, and adding stable frames. A smoothed histogram of the whole recording is plotted on the right vertical axis. Only a portion of the melograph is shown in the zoomed view, where frequency information is presented as intervals with respect to the tonic. The tonic, automatically found as in Bozkurt (2008), is open to manual correction.

pruning by manual selection. Various options for automatic pruning can also be implemented. For example, for automatic pruning, the pitch histogram can be used to discard all segments with pitches that lie outside of a range of frequencies around histogram peaks. Another pruning method would be to throw out short segments if longer segments exist that have very close f_0. A decision based on f_0 dynamic range (i.e., preferring the frame with the lowest dynamic range among candidates for a tone) is another possibility. In Figure 3, a user interface designed for manual pruning is presented, with example outputs available online at akademik.bahcesehir.edu.tr/~bbozkurt/tuner. Automatic pruning is planned to be part of our future research and will not be further detailed here.

The analysis part of the tuner is designed as a user interface (shown in Figure 3) to perform manual pruning of segments, as well as to set the minimum length and maximum deviation parameters defined previously. For manual segment selection, the interface provides the user with the frames as labeled segments on melographs (Cohen 1964), where the user can then click, listen, and make a selection. In addition, the user can specify stable frames that were not detected by the system. This is done by observing the melograph, manually segmenting and listening to specific portions, and then specifying the stable region with a button click.

It may seem tedious to perform manual pruning, but it would have to be done only once for a given recording, after which users can exchange their sets and get help from their masters. For a 3-minute recording, the author's experience is that the time needed for manual pruning is around 10 minutes. For traditional music, we think it would be beneficial to include the master in this process. This way, the tuner becomes a tool suited for oral traditions, where the master gives the pupil a set of sounds of the master's choosing, rather than having the tuner provide certain frequency values or some "globally used" sound signals. The act of manual pruning would also help the learning process, as it requires listening, observing (histograms, melographs), and making choices. For users who are too inexperienced for such a process, the developer or experienced users can share sets directly or through a Web site. For TMMT we have created sets to be shared on the Web. Examples are provided for makams Saba and Hüseyni online at akademik.bahcesehir.edu.tr/~bbozkurt/tuner. For other music styles, a similar effort by advanced users would help other inexperienced users.

Creating Loopable Versions of Stable Frames

Once the set of stable reference frames is available, loopable versions must be created for continuous audio playback during the tuning process. If the frame itself is looped (even after a phase correction to guarantee signal continuity in looping), the inherent variations in the signal would be periodically repeated, making tuning more difficult. The procedure described here reduces the periodicity induced by looping.

The loopable versions are obtained in the following steps: filtering and computing the nth power in the frequency domain, removing the time-domain envelope, and windowing in the time domain to guarantee phase continuity. In the following paragraphs we present the details for these steps.

The procedure starts with computing the Fourier transform of the frame with the same number of data points in both the time and frequency domains. In the frequency domain, first a filter is applied to remove all components except the fundamental. Because the rough f_0 value is available for the frame, this filter is simply composed of multiplication with a band-pass filter (a rectangle with an ideal frequency response) with cut-off frequencies set to $\omega_L = 0.5 f_0$, $\omega_H = 1.5 f_0$. Then, a point-wise power is obtained (preferably of degree higher than $n = 5$) that boosts the fundamental (highest-energy) component in the signal. For a discrete-time signal $x[n]$ with length N, $X(k)$ corresponds to the discrete Fourier transform of $x[n]$. $X_f(k)$ is the filtered version and $X'_f(k)$ is the version after the power operation (with real order p of typical value higher than 5) and $W(k)$ is the filter response.

$$X(k) = DFT\{x[n]\},$$
$$n = 0, 1, ..N - 1, \quad k = 0, \quad 1, ..N - 1$$

$$X_f(k) = X(k)W(k),$$
$$W(k) = \begin{cases} 1, & k_L < k < k_H, N - k_H < k < N - k_L \\ 0, & otherwise \end{cases}$$

$$X'_f(k) = X_f(k)^p, \quad k = 0, 1, ..N - 1$$
$$x'[n] = IDFT\{X'_f(k)\}$$

The signal obtained, $x'[n]$, is very close to a sinusoid at the fundamental frequency of the original frame. A time-domain envelope is still observed, however, due to the inherent variations in the frame. This also has to be removed because, when looped, $x'[n]$ may contain a beat-like component due to low-frequency variations. The time-domain envelope can easily be detected by the Hilbert transform as the amplitude of the analytical signal, and can be removed by a simple point-wise division operation. Finally, amplitude normalization is applied.

An example of these procedures follows. The corresponding sound files are available online at akademik.bahcesehir.edu.tr/~bbozkurt/tuner. A short ney improvisation, in makam Saba and recorded by Niyazi Sayın, one of the living masters of TMMT, is made available. (The ney is an end-blown flute frequently used in TMMT.) The pitch histogram, prepared as described in Bozkurt (2008), is presented in Figure 4. This histogram shows that most of the tone frequencies do not match either twelve-tone equal temperament (12-TET) or the official standard tuning system for TMMT (Arel 1968).

Figure 5 shows the automatically segmented stable frames with maximum deviation of 25 cents and minimum duration of 250 msec, for the first 24 seconds of the recording. We observe that several instances of the same pitch are labeled. For one of the frames, the intermediate signals obtained are presented in Figure 6.

The procedure described is highly efficient (although not perfect) in obtaining loopable versions that carry mainly the fundamental harmonic and do not include perceivable variation from looping. Still, it is beneficial to avoid frames with large vibrato during manual pruning. In some cases, all variations in the frame cannot be removed if the variations are too large or the frame contains more than one harmonic very close in frequency. One example is the recording of double strings, where the two strings do not exactly match in frequency or the pressure applied on both of the strings on the fret by the player's finger is not equal; another example is some ney sounds, because the instrument's acoustic tube surface is an unprocessed, natural material. Indeed, in Figure 6, some components are observed

Figure 4. Pitch histogram of Saba taksim by Niyazi Sayın. Vertical lines indicate 12-TET tones and circles indicate tones for makam Saba in TMMT theory. The x-axis shows frequency on a logarithmic sale, using two different units: Hc, corresponding to the Holderian-Mercator comma (1/53 of an octave), and cents.

Figure 5. Melograph with automatically segmented stable frames indicated with bold lines.

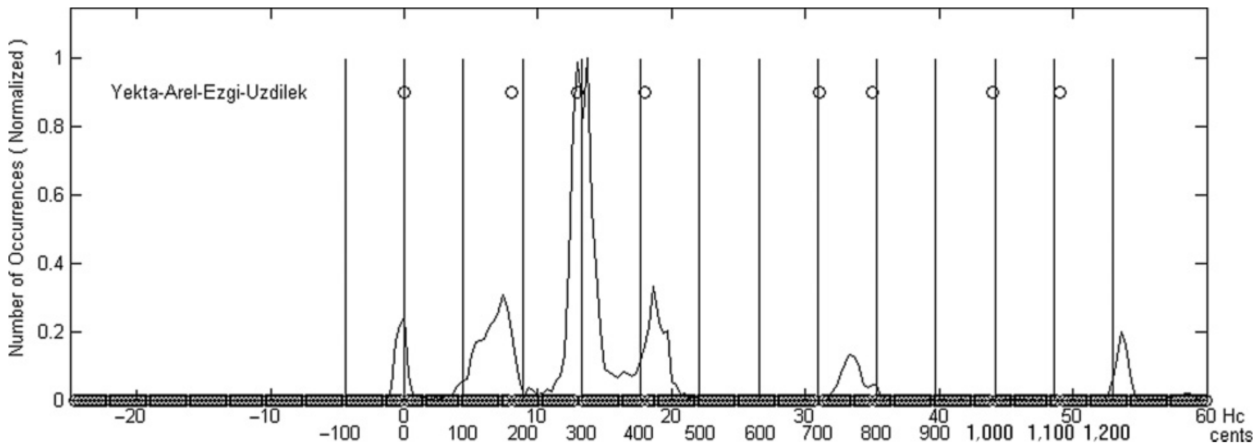


Figure 4

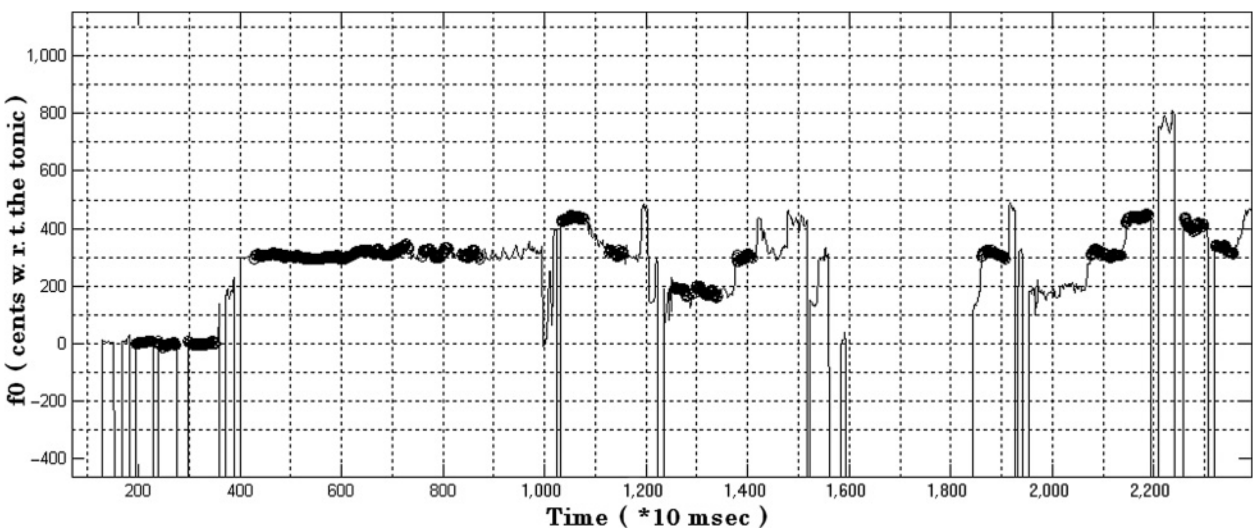


Figure 5

at $k = 20$ and $k = 400$ (marked with rectangles on the figure), which are not close to integer multiples of the main harmonic. Some stable frames obtained (by manual pruning of automatically segmented frames) from two recordings are presented on the Web site together with the loopable versions. Among these, the second frame for Niyazi Sayın's recording is an example where variation can still be heard in the looped version.

The Application Part

The application part of the software is designed to provide audio and visual feedback to help the user match his or her instrument's sound with the reference frames extracted by the offline module.

The audio feedback is played through a headphone output as a weighted sum of the looped reference sound and the actual input signal.

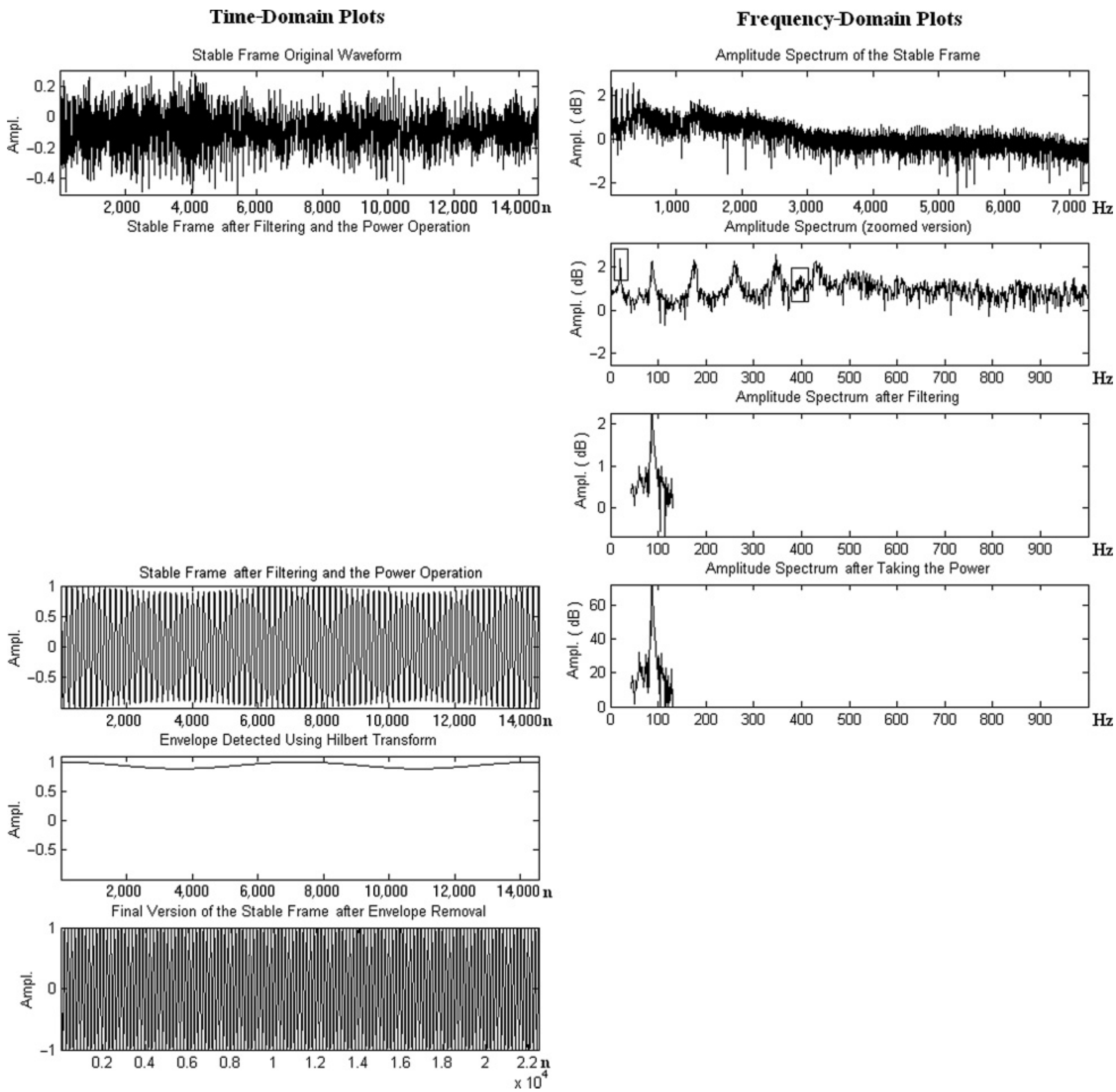


Figure 6. Example for intermediate signals in the analysis, all at a sampling rate of 16,000 Hz. Row 1, Column 1: the automatically segmented stable frame. Row 1, Column 2: amplitude spectrum of the stable frame. Row 2, Column 2: amplitude spectrum (zoomed version). Row 3, Column 2: amplitude spectrum (filtered version). Row 4, Column 2: 32nd power of the filtered amplitude spectrum. Row 4, Column 1: time-domain signal after filtering and power computing in the frequency domain. Row 5, Column 1: envelope computed using Hilbert transform. Row 6, Column 1: final loopable version after removal of the time-domain envelope.

The weight is adjusted by the user to keep one sound from dominating another. The user also has the choice of listening to the original, stable frames.

The visual display is composed of plotting auto-difference functions of the two signals (the loopable version of the reference signal and the input signal). An auto-difference function very clearly reflects the pitch-period information at the location of the peaks. Hence, by matching peaks, the user can match the frequencies of two signals without estimating any frequency value.

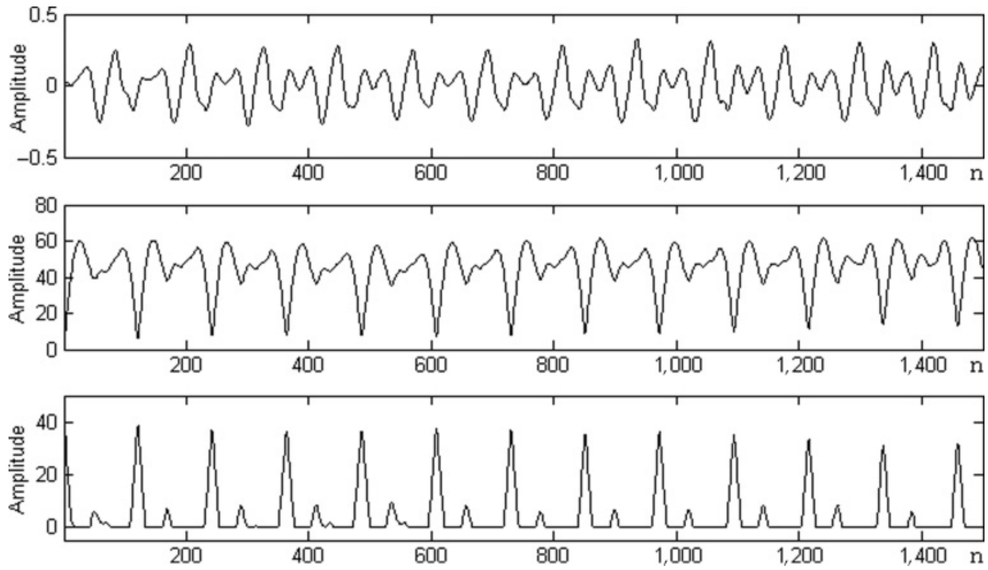


Figure 7

Figure 7. Kemençe signal (top), auto-difference function (middle), and shaped auto-difference function (bottom).

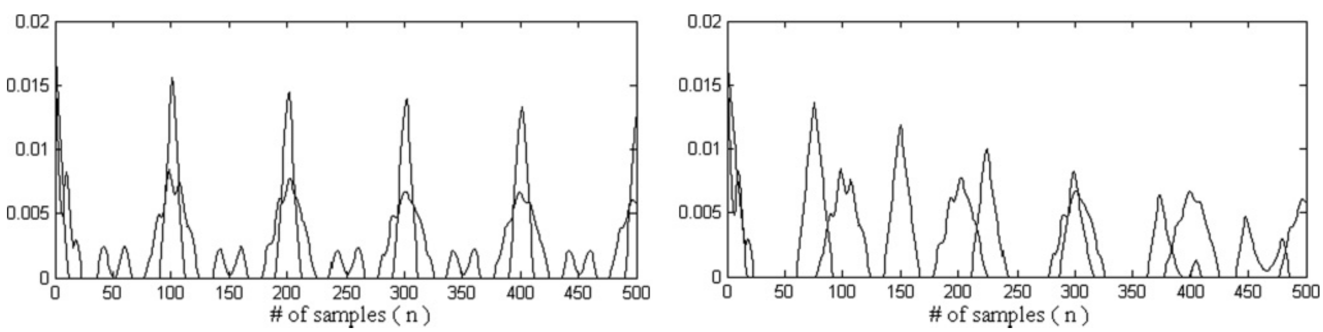


Figure 8

Figure 8. Examples of auto-difference function comparison. All signals are sampled at 22,000 Hz. Left: Re-shaped auto-difference functions of singing at 220 Hz and kemençe at 220 Hz. Right: Conditioned auto-difference functions of singing at 220 Hz and kemençe at 300 Hz (higher frequency implies shorter period). The auto-difference function of the singing sample is exactly the same on both figures. The only change is the frequency of the kemençe sample.

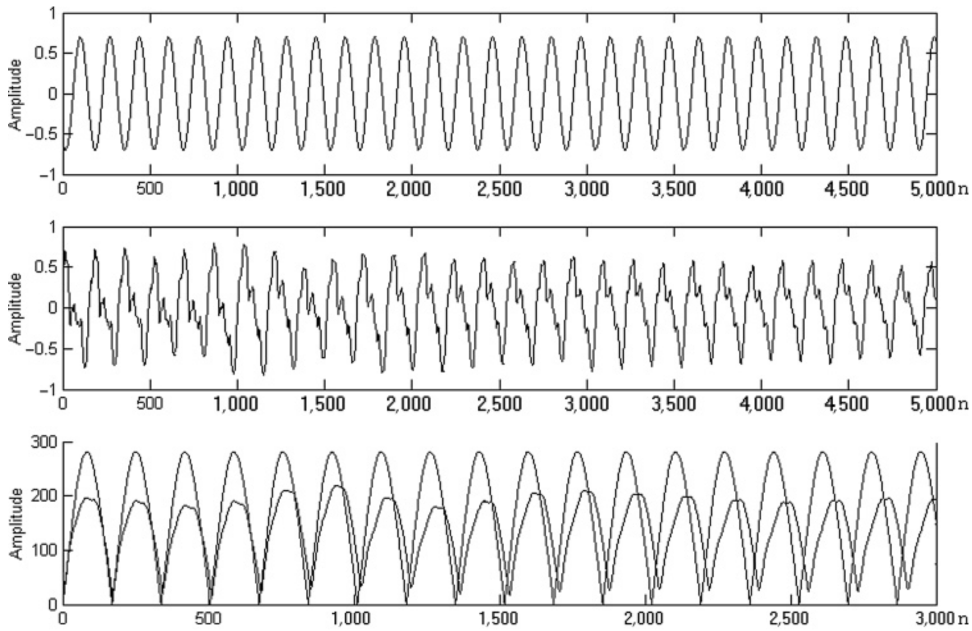
The auto-difference function, $d[n]$, of a discrete-time signal, $x[n]$, is computed using the following formula, as implemented in YIN (de Cheveigné and Kawahara 2002).

$$d[n] = \sum_{k=1}^{K} \left| x[k] - x[k+n] \right|$$

To shape the auto-difference function to enhance peaks, the following steps are applied: polarity inversion, subtraction of the mean, and half-wave rectification. In Figure 7, we present an example sound signal of a kemençe (a traditional instrument used in TMMT), its auto-difference function, and the shaped auto-difference function.

In YIN, the auto-difference function was used to estimate the pitch period. As this function reveals the pitch clearly, it can also be directly used as a visual feedback of the pitch of a sound. This makes possible an approach to tuning without tracking the fundamental frequency. As an example, in Figure 8 we present signals from two different instruments:

Figure 9. A loopable version reference frame (top), a slightly out-of-tune guitar signal (middle), and the auto-difference functions of the two signals (bottom).



voice and kemençe. Although the timbres are quite different, peak comparison gives obvious information about the degree of correspondence between the frequencies.

In the tuner design, loopable versions of the reference frames show very clear peaks, as would be expected from a sinusoid-like signal. One such example, together with a guitar signal that is slightly out of tune with the reference, is provided in Figure 9. Through practice with the tuner, it is observed that a zoomed-in display of a region of several high-indexed cycles is preferable to displaying the first few cycles of the auto-difference function (i.e., toward the right-hand side in Figure 9), because the frequency deviation can be more easily observed. A video demonstration has been created during the tuning of open guitar strings to one of the stable frames found from the ney example (in Figures 4 and 5). In this demonstration, the view is zoomed in to higher harmonics owing to the previously mentioned observation. We believe the video provides a good demonstration of how frequency difference can be viewed on auto-difference functions. Zooming the view in and out, as is common with handheld devices, can easily be made user-controllable on the tuner display to further facilitate usage.

Testing the System's Usability

This article proposes a new approach to tuning and to tuner system architecture. A complete product is not yet available; hence, complete testing is out of our scope. User feedback, however, provides insight into the usability of such a system.

A prototype (lacking the real-time visual feedback) has been implemented in MATLAB and has been presented to master musicians of TMMT on various occasions. There, only the audio feedback functionality is used for actual tuning. We will now summarize the feedback obtained from these musicians about the system's potential use in the context of TMMT.

Most of the masters considered the prototype more as a research tool than as an actual, usable device. Two important reasons were that the tuner needs user input (providing a recording and making choices for the references) and that it runs on a computer rather than a simple dedicated hand-held device. They all stated that it is an important research goal to experiment with different tunings used by past master musicians. About one-third of the masters stated they would like to own the system on a hand-held device and test it in their ensembles.

As a case study, we decided to perform a hands-on test with a master kanun player, because the device would be very beneficial to kanun players (who, as described previously, need to tune 24 to 26 triplets of strings). The kanun is a difficult-to-tune instrument, both because of the number of strings and also because it is a fixed-pitched instrument in a microtonal music setting.

During half of a day, the prototype device was tested (with the help of the author) by Reha Sağbaş, a respected kanun player with 30 years of performance experience in Turkish Radio Television ensembles. Reha Sağbaş has contributed to many albums as a musician or art director, has received world music awards, and has given concerts in many countries. He is well known in professional music circles for his special interest on the tuning problem in TMMT.

For testing the system, two recordings of Tanburi Cemil Bey (Cemil 1994) were used for tuning the kanun of Reha Sağbaş. Then, Reha Sağbaş was asked to play an improvisation for each and comment on the result. Tanburi Cemil Bey is often cited as the most influential musician of 20th-century TMMT. His style has been imitated by many famous musicians. Reha Sağbaş stated that he considers Tanburi Cemil Bey's recordings, especially the solos, the best representative examples of 19th- and 20th-century makam music. Even though he greatly admires his musicianship, this is the first time he could actually tune as in the recording of Tanburi Cemil Bey, because of the previous unavailability of an adequate tuning device. Before the experiment, he stated that the way he tunes his instrument is based on the iterative process of playing and tuning with other musicians. In these cases the reference varies over time. During the last 20 years, there has been a shift towards equal temperament due to the increasing use of western-music tuners. One of his unreached goals is to develop a method or system to facilitate proper tuning of orchestra instruments, stopping this shift towards equal temperament.

The reference recordings, and the resultant solo improvisation recordings, are available online at akademik.bahcesehir.edu.tr/~bbozkurt/tuner. The first recording was in makam Rast. The solo improvisation by Reha Sağbaş is composed of two parts. In the first part, he plays phrases in the main mode. In the second part he tests close as well as distant modulations to observe the effect of the tuning on the other modes. Both sections end with a short commentary in Turkish. The short summary of his comments is as follows: He states that such a tuning would be considered as mistuning in the ensembles he plays in, and that this is the first time he has played using such a tuning in his entire career. To exemplify, while explaining he plays a major second (sol-la interval; at 15:33 in the recording) and states that nobody would tune a major second that way; yet, in the melodic progression, it sounds very restful. It is surprising that no one tunes that way, because Tanburi Cemil has been a major influence on contemporary musicians. The second recording used is in makam Suzidil. Reha Sağbaş's instrument was re-tuned and he played a solo improvisation. This time, he was more familiar with the process and more comfortable in applying the tuning and performing an improvisation.

His overall conclusion is that the system is both a tuning and a research tool. The user interface is confusing, therefore older masters would not be inclined to use it. He thinks that a collective effort should be put to obtain reference segments for the most important musicians of the last century and that these should be shared on the Web. Then, the application part alone can function as the tuner, and he thinks that this would be an important contribution to the future of TMMT.

The visual feedback functionality has only been tested by the author on various files by offline creation of videos. For actual testing with users, real-time signal processing and viewing is needed. This has not yet been implemented by the author. This part of the design is tested in a proof-of-concept scenario using plots and videos created offline. Three videos are available on the Web page previously mentioned. The first one demonstrates the view during the tuning process. For creation of the video, first, the sound of the instrument to be tuned is recorded while using only the audio feedback of the tuner. Then the video is created, offline, using the procedure proposed for the visual feedback creation (implemented in MATLAB), taking as input the reference looped signal and recorded instrument sound. The second and third videos differ only by using a synthetic signal, obtained from YIN f_0 estimates, to compare the system output with YIN output. In these two videos we observe that YIN and the proposed viewing match most of the time, where only in the last sound, the YIN results include some octave errors (halving in one case and doubling in the other).

Conclusion

This article has proposed a new tuner suited to traditional music. A prototype has been implemented by the author and some proof-of-concept results are available online at akademik.bahcesehir.edu.tr/∼bbozkurt/tuner.

For music of oral traditions, tuning is often a controversial issue, unlike music with a commonly accepted tuning theory. The use of existing technology with preset theoretical frequencies becomes debatable, even if in many cases there is no other way to be able to play in tune together. In this case, it is obvious that a risk has been taken in that the technology may have transformed the tuning system of the particular tradition.

The proposed tuner system involves an analysis part and an application part, both of which involve several signal-processing problems. Although this article has proposed a solution for each problem, all are open to further improvements. One such improvement is the design of an efficient automatic pruning system for beginner-level musicians who would have difficulty with manual pruning.

We believe that the tuner would be particularly useful for instrument makers of fretted or other fixed-tone instruments. In oral traditions, most of the time the instrument-making process is a template-copying operation. For example, in TMMT the tanbur is constructed using templates for fret locations for a few fixed neck sizes. With a newly built instrument, the musician still has to adjust by ear the location of the bridge and some of the frets. The tuner would provide instrument makers with the flexibility to give up using templates and build new instruments with different sizes or shapes while keeping the desired pitches.

Limited tests of the tuner have been performed in the context of TMMT. The tuner prototype has been used to tune instruments of a few masters using mainly the audio feedback of the application part. Demo recordings and videos obtained during that process are available. Our overall conclusion is that the analysis system seems too sophisticated to musicians of traditional music, and a simple and user-friendly product is needed for actual use. They consider the system more as a useful research tool. The real potential of the tuner can be tested once a final product is available. For that aim, implementing the application part as a mobile-phone application appears to be a good choice, which is among our future goals. The analysis part can then be used by interested young musicians and researchers, and reference frames can be shared on the Web for other users.

Learning a traditional music that has no commonly accepted tuning system (such as TMMT) makes playing correct pitches very difficult. The proposed tuner is potentially useful for learning and can be used as part of interactive tutoring systems. This is also part of our future research.

References

Arel, H. S. 1968. *Türk Musikisi Nazariyatı Dersleri*. Istanbul: Hüsnütabiat Matbaası. Reprinted in Onur Akdogu, ed. 1993. *Kültür Bakanlığı Yayınları.* 2nd ed., vol. 1347. Ankara: Repromat Basımevi.

Bello, J. P., et al. 2005. "A Tutorial on Onset Detection in Music Signals." *Institute of Electrical and Electronics Engineers Transactions Speech Audio Process*ing 13(5):1035–1047.

Bozkurt, B. 2008. "An Automatic Pitch Analysis Method for Turkish Maqam Music." *Journal of New Music Research* 37(1):1–13.

Bozkurt, B., et al. 2009. "Weighing Diverse Theoretical Models on Turkish Maqam Music Against Pitch Measurements: A Comparison of Peaks Automatically Derived from Frequency Histograms with Proposed Scale Tones." *Journal of New Music Research* 38(1):45–70.

Camacho, A. 2007. "SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech and Music." PhD dissertation, University of Florida. Available online at www.cise.ufl.edu/∼acamacho/publications/dissertation.pdf. Accessed April 2012.

Cemil. 1994. *Tanburi Cemil Bey*. New York: Traditional Crossroads CD 4264. Remastered from the Orfeon 10498 original 78 RPM record, 1910–1914.

de Cheveigné, A., and H. Kawahara. 2002. "YIN, a Fundamental Frequency Estimator for Speech and Music." *Journal of the Acoustical Society of America* 111(4):1917–1930.

Cohen, D. C. 1964. "An Investigation into the Tonal Structure of the Maqamat." *Journal of the International Folk Music Council* 16:102–106.

Hess, W. 1983. *Pitch Determination of Speech Signals: Algorithms and Devices.* Berlin: Springer.

Holzapfel, A., et al. 2010. "Three Dimensions of Pitched Instrument Onset Detection." *IEEE Transactions on Audio, Speech and Language Processing* 18(6):1517–1527.

Moore, B., B. Glasberg, and T. Bear. 1997. "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness." *Journal of the Audio Engineering Society* 45(4):224–239.

Schloss, A. W. 1985. *On the Automatic Transcription of Percussive Music from Acoustic Signal to High-Level Analysis.* PhD dissertation, Stanford University, Department of Hearing and Speech.

Six, J., and O. Cornelis. 2011. "Tarsos—A Platform to Explore Pitch Scales in Non-Western and Western Music." In *Proceedings of the International Society for Music Information Retrieval*. Available online at ismir2011.ismir.net/papers/OS2-4.pdf. Accessed April 2012.

Tidhar, D., M. Mauch, and S. Dixon. 2010. "High Precision Frequency Estimation for Harpsichord Tuning Classification." In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 61–64. Available online at http://www.eecs.qmul.ac.uk/∼simond/pub/2010/Tidhar-Mauch-Dixon-ICASSP-2010.pdf. Accessed April 2012.

Yarman, O. 2008. "79-Tone Tuning & Theory for Turkish Maqam Music." PhD dissertation, Istanbul Technical University, Social Sciences Institute.

Zhou, R., and J. D. Reiss. 2007. "Music Onset Detection Combining Energy Based and Pitch-Based Approaches" (MIREX Onset Detection Contest.) In *Proceedings of the International Society for Music Information Retrieval*. Available online at www-student.elec.qmul.ac.uk/people/josh/documents/ZhouReiss-MIREX2007-onsetdetection.pdf. Accessed April 2012.