ARCHIVES OF ACOUSTICS
33, 4 (Supplement), 99–104 (2008)

AUTOMATIC DETECTION AND CORRECTION OF DETUNED SINGING SYSTEM FOR USE WITH QUERY-BY-HUMMING APPLICATIONS

Michał LECH

Gdańsk University of Technology
Multimedia Systems Department
Narutowicza 11/12, 80-952 Gdańsk, Poland
e-mail: mlech@sound.eti.pg.gda.pl

(*received June 15, 2008; accepted October 31, 2008*)

The aim of the paper is to present an idea of using the automatic detection and correction of detuned singing as a subsystem in query-by-humming (QBH) applications. The common approach to searching for a requested song basing on the melody retrieved from hummed pattern usually employs the so-called Parsons code or melody contour. In such a case information about sound pitch is discarded. It was thought out that an additional module added to the QBH system indicating notes which were sung out of tune and correcting them might be useful. For this purpose two fundamental frequency detection algorithms, i.e. the fast autocorrelation and HPS (Harmonic Product Spectrum), and two pitch shifting algorithms, i.e. the modified phase vocoder and PSOLA (Pitch-Synchronous Overlap-Add) are chosen and examined. Four possible combinations of the algorithms are reviewed in the context of correctness of the fundamental frequency detection and pitch shifting. Basing on the results, the sub-system for automatic detection and correction of detuned singing for use with QBH applications is implemented. In addition, listening tests and objective measurements of the obtained pitch correction are performed. Conclusions are drawn and proposals of further improvements are provided.

Keywords: pitch shifting, melody retrieval, query-by-humming, correcting pitch.

1. Introduction

Since the idea of query-by-humming (QBH) systems was born many methods of melody retrieval have been developed which show that the problem of finding the proper melody among thousands of possibilities is not trivial [3, 6, 9]. Indeed, to provide reliable melody finding, the mechanism must be resistant to such features, regarding melody hummed by the user, as for example: pitch transpositions, tempo changes, note insertions and deletions [6, 11]. Considering existing methods and algorithms trying to deal with the problem, such as dynamic time-warping approaches [5], Markov model techniques [10] or dynamic programming [9], one can notice that the great effort is put on the melody retrieval mechanism basing on the unaltered user's singing sample.

Conversely, none of the methods attach importance to improvement of the hummed melody. The methods base on the fact that not many people can sing in tune and assume that the given singing sample cannot be corrected by them. In fact, even experienced musicians can have problems with singing in tune but listening to their recorded voice they can without any problems indicate excerpts in which notes were not properly sung. Similar situation can be mentioned while considering amateur singing. Basing on this observation, in the paper the system of pitch correction was proposed as the additional module for QBH applications, which enables the user to correct pitches according to his memorized form of the original melody that he is searching for.

2. Typical melody representation

QBH systems for retrieving purposes usually employ melody contour of the user's input, which is a simple notation used to identify a piece of music through melodic motion. The simplest way is to use three characters describing pitch changes, i.e. U (up) for pitch upper than the previous one, D (down) for pitch lower than the previous one, and R (repeat) or S (same) for repeated pitch. Describing melody in such manner is known as Parsons code [8]. In such an approach different melody contours can have identical Parsons code representation [9]. Melody contour based on Parsons code also discards information about rhythm. Such a melody representation can lead to inefficient searching [1,9]. More precise five step melody contour representation, which additionally includes rhythmical information, is provided by the international MPEG-7 standard. Melodic contour intervals, presented in Tab. 1, are contained in the field *Contour* of the MPEG-7 output format *MelodyContourType*. The second field of the format, called *Beat*, contains numbers of beats where contour changes occur (truncated to whole beats) [1, 2].

Table 1. MPEG-7 melody contour values for 5 step representation.

Contour value	Pitch change expressed in cents
-2	$\Delta f_c \leq -250$
-1	$-50 \leq \Delta f_c < -250$
0	$-50 < \Delta f_c < 50$
1	$50 \leq \Delta f_c < 250$
2	$\Delta f_c \geq 250$

3. Pitch correction module

Representing melody even with a more detailed MPEG-7 melody contour can still cause some inaccuracies [7]. For instance, apart from singing out of tune, user can use drastically different intervals from these contained in the original melody. It is quite common that some intervals are favored by amateurs, although they sound worse in the sequence of notes than the original ones. The reason for this is the fact that they are easier to sing or hum. Thus, two similarly sounding to many persons melodies can have highly different melody contours. As a solution QBH systems use many methods of approximate pattern matching but the more user's melody contour differs from the original one the more melodies given to the user as the searching result [9].

Listening to self-hummed melodies user can be conscious of their inaccuracies but without some knowledge of tonality rules or ability to hum in steady key he is not able to improve them. Therefore, a tool enabling users to correct their melodies by pitch shifting their notes, resulting in an additional module to QBH system can be considered. The structure of the module is presented in Fig. 1.

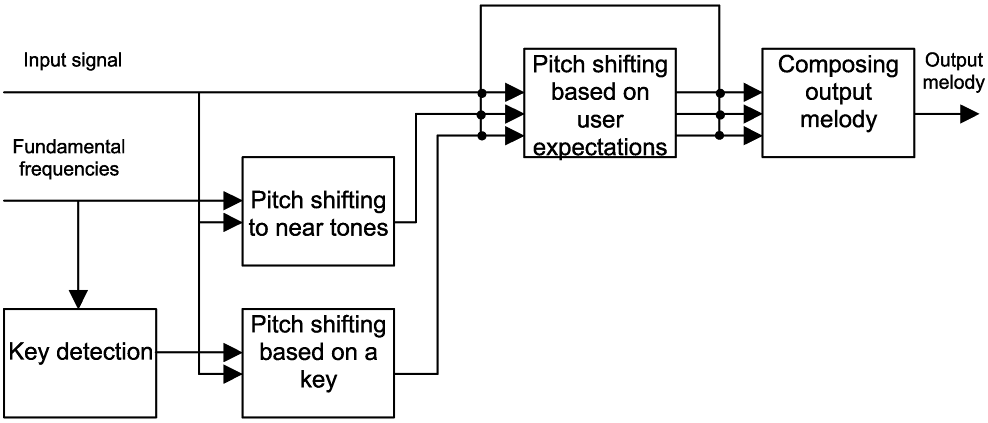
Input signal
Fundamental frequencies
Key detection
Pitch shifting to near tones
Pitch shifting based on a key
Pitch shifting based on user expectations
Composing output melody
Output melody

Fig. 1. Structure of the pitch shifting module for QBH systems.

At the first stage a key detection is performed basing on the relationship between notes with regard to music harmony [12, 13]. In the next step two copies of the signal are pitch shifted. Pitches of the first copy are increased or decreased to match accurately the nearest tone. The second copy is pitch shifted according to the detected key in such manner that the nearest tone from the key is achieved for the particular note with the additional assumption that if the previous note is hummed lower than the considered one, then the pitch of the considered note is increased, and otherwise decreased. The user has a possibility to listen to each signal and decide which one is the most similar to the original melody. Each signal pitches are displayed as blocks positioned in the piano-roll. Each pitch can be individually adjusted by selecting its corresponding block and changing its position. For the obtained, corrected signals the user can change pitch with semi-tone step. For the pitch corrected signal with regard to the detected key, also pitch changes with the interval acceptable in the key are suggested by the system. After adjusting the signals the user has the possibility to choose excerpts of each signal, among which the corrected notes reflect the original ones according to his memory of the original melody. The excerpt selection is performed by selecting particular notes. Basing on the chosen excerpts the output signal is composed which is the subject to melody representation and melody retrieval QBH system phases.

The placement of the pitch shifting module in the QBH system is shown in Fig. 2. After finishing the tasks of the module, no signal operations of fundamental frequency detection are needed to define pitches, as during pitch shifting phases, for each note symbolic pitch description is derived from the knowledge of original fundamental frequencies, pitches defined during key detection and shifting intervals chosen by the system or a user. Considering melody representation, MPEG-7 melody contour can be used and additionally searching based on pitch contour with the information on exact intervals can be performed. The approach proposed bases on the fact that the signal corrected by the user using described method is closer to the searched melody than the original unaltered hummed pattern and more detailed melody retrieval rules can be applied.

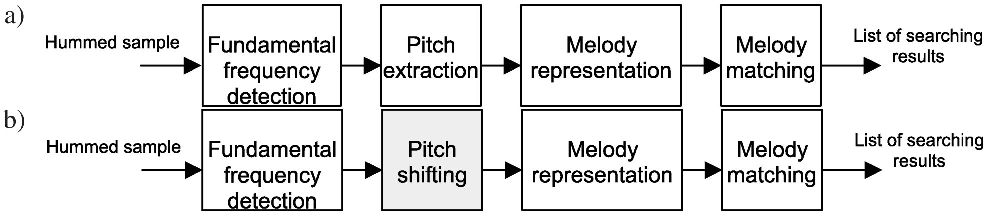

Fig. 2. Basic QBH system architecture (a) and the architecture of the QBH system with pitch shifting module added (b).

4. Research on algorithms

Objective measurements of two fundamental frequency detection algorithms, i.e. fast autocorrelation (time-domain method) and HPS (*Harmonic Product Spectrum*; frequency-domain method), and two pitch correction algorithms, i.e. PSOLA (*Pitch-Synchronous Overlap-Add*; time-domain method) and modified phase vocoder (frequency-domain method) were performed together with listening tests.

Tests were performed using male and female with glissando articulation singing samples and sequences of single notes. Various frame lengths were used. The correction based on increasing the first tone of the glissando and preserving it for the whole duration of the sample. It was assumed that the proper correction was such that the resulted pitch equaled the reference pitch and quality was subjectively rated as level of general similarity in sound with the original signal.

Analysis of the results has showed that the correction effectiveness depends on the particular fundamental frequency detection algorithm. Using the autocorrelation algorithm with a long frame resulted in skipping tones of a short duration (shorter than the frame length). Such a problem did not exist using the HPS algorithm as it does not operate on the time-domain form of a signal.

The research on a quality of corrected signals depending on length of the used frame showed that for the PSOLA algorithm the shorter frame used, the more audible distortion or flutter to the sound. For the modified phase vocoder there was no correlation between frame length and sound quality observed but a negative effect on formants resulting in unnatural sound was noticed.

5. Implementation and validation

Basing on the results of the research described in the previous Section it was concluded that the optimum choice for the correction of detuned singing are HPS and PSOLA algorithms. The system was implemented in JAVA, as it provides many, free sound libraries.

The pitch correction provided by the system was validated using male singing sample consisted of notes H3 to G4 sung in sequence and female and male glissando articulations used previously for testing Matlab algorithms. For the sequence of tones four MIDI test patterns were used. The first two patterns contained sequences increased and decreased by whole tone. The third pattern consisted of sung notes, therefore its aim was to level each out of tune note. Tones of the last pattern were determined by random number generator giving numbers from 59 (note H3 MIDI code) to 67 (note G4 MIDI code). For male and female glissando articulation three patterns were prepared. The first pattern consisted of the note beginning the glissando increased by a whole tone, the second one — the note with which the glissando begun and the third one — the note beginning the glissando decreased by a whole tone.

Analyzing sequences pitch shifted by a whole tone it was observed that three last tones were not shifted correctly (fundamental frequency detection effectiveness equaled respectively 5.3%, 0.0% and 1.7% for sample containing notes decreased by a whole tone and 71.4%, 70.6%, 5.2% for sample consisting of notes increased by a whole tone). For other notes the average fundamental frequency detection effectiveness equaled 81%. When using randomly generated MIDI pattern, although pitch is shifted correctly, quality of resulting sound is very low. For glissando articulations pitch shifting was done properly. The analysis of the singing sample allows concluding that the problems were caused by voice articulation. Three last notes were sung with much greater attack than the others.

6. Conclusions

The aim of the paper was to propose a pitch correction module for use with query-by-humming applications. Using such a module user is able to correct his singing sample or hear various modifications of the melody he sang in case he is not sure if it is accurate in comparison with an original vocal line. The subsystem can be especially useful for persons who have an ear for music but have problems with singing in tune. Possessing a singing sample which corresponds to the original melody in a subjective judgment of a user gives a possibility of creating more strict melody retrieving rules which in consequence lead to more efficient searching. For further improvements one could consider variable frame length automatic selection module, presented in work by [4] and also a module of a rhythm correction.

References

[1] BATKE J., EISENBARG G., WEISHAUPT P., SIKORA T., *A Query by Humming system using MPEG-7 Descriptors*, AES 116th Convention Paper 6137, Berlin 2004.

[2] BATKE J., EISENBARG G., WEISHAUPT P., SIKORA T., *Evaluation of Distance Measures for MPEG-7 Melody Contours*, IEEE, 131–134, (2004).

[3] KOSTEK B., *Applying computational intelligence to musical acoustics*, Archives of Acoustics, **32**, 3, 617–629 (2008).

[4] LECH M., KOSTEK B., *A system for automatic detection and correction of detuned singing*, lay-language version of paper 1462, Acoustics'08, Paris 2008.

[5] MAZZONI D., *Melody matching directly from audio*, Proceedings of International Symposium on Music Information Retrieval, 2001.

[6] MEEK C., BIRMINGHAM W., *The dangers of parsimony in query-by-hamming applications*,

[7] GÓMEZ E., BOUYON F., HERRERA P., AMATRIAIN X., *Using and enhancing the current mpeg-7 standard for a music content processing tool*, Proc. of the 114th AES Convention, Amsterdam, March 2003.

[8] PECHELT L., TYPKE R., *An interface for melody input. ACM Transactions on Computer-Human Interaction*, **8**, 2, 133–149 (2001).

[9] RHO S., HWANG E., *Query adaptive melody retrieval system*, The Journal of Systems and Software, **79**, 43–56 (2006).

[10] SORSA T., *Melodic resolution in music retrieval*, Proceedings of International Symposium on Music Information Retrieval, 2001.

[11] WAKEFIELD G., *Vocal Pedagogy and Pedagogical Voices*, Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, 6–9 July, 2003.

[12] ZENZ V., RAUBER A., *Automatic chord detection incorporating beat and key detection*, conference paper.

[13] ZHU Y., KANKANHALLI M., GAO S., *Music Key detection for musical audio*, Proceedings of the 11th International Multimedia Modelling Conference (MMM'05), 1550–5502/05, 2005.