IDENTIFICATION OF MUSICAL CHORDS USING CONSTANT-Q SPECTRA

S. Hamid Nawab, Salma Abu Ayyash, and Robert Wotiz

ECE Department, Boston University
Boston, MA 02215
email: hamid@bu.edu

ABSTRACT

We present an approach to the extraction of frequencies corresponding to chords in western polyphonic music. In the first phase of this approach constant-Q spectral analysis directly provides the features from which the fundamental frequencies for 43 of the 57 possible categories of chords can be extracted without ambiguity. Each remaining chord category has a potential ambiguity associated with it because of frequency resolution problems. The second phase of our approach is designed to address such ambiguities. A software implementation of our approach was used successfully to validate its performance on a representative set of polyphonic musical signals.

1. INTRODUCTION

Polyphonic music is produced by one or more musical instruments playing several co-occurring melodies. A primary set of inter-note constraints specify the valid relationships between the fundamental frequencies (F0's) of simultaneously played notes in western polyphonic music [1, 2]. For example, combinations of simultaneous notes, which are called chords, can have up to four different constituent notes with distinct F0's. The total number of permissible chords is 57 as dictated by the allowable ratios of the F0's of the constituent notes [2].

In order to identify the chords represented in a polyphonic music signal, we consider the problem of extracting F0's from an analysis of time-dependent signal spectra. A viable approach to spectral analysis of *monophonic* music is the Constant-Q (CQ) transform [3]. Furthermore, with a $Q \geq 17$, the CQ transform adequately deals with the time-dependent frequency characteristics [1] of such signals [2]. For polyphonic music signals, a CQ spectrum is not guaranteed to contain peaks corresponding to constituent fundamentals. To further appreciate this result, we note that a signal component at f_1 will give rise to a spectral peak at f_1 provided there is no other signal component at any frequency f_2 such that:

$$|f_1 - f_2| < \frac{f_1}{Q}. \tag{1}$$

Otherwise, the component at f_1 gives rise to a peak at f_1' where:

$$f_1 - \frac{f_1}{2Q} \leq f_1' \leq f_1 + \frac{f_1}{2Q}. \tag{2}$$

This work was sponsored in part by USAF Rome Laboratory under Contract No.F30602-95-C-0204.

[1] such as a vibrato of up to 10Hz, a tremolo of up to 10Hz and a vibrato depth of up to 1.5%

Combining the results in (1) and (2), we conclude that a component at frequency f_2 will render the component at f_1 unresolvable if:

$$\frac{Q}{Q+1} f_2 \leq f_1 \leq \frac{Q}{Q-1} f_2. \tag{3}$$

An analysis of the permissible chords in western polyphonic music shows that 14 out of the 57 cases have a resolvability problem. For example, suppose that the lowest fundamental in a chord is at f_0. Let us consider a chord composed of 4 notes with respective F0's of f_0, $2^{10/12} f_0$, $2^{19/12} f_0$ and $2^{28/12} f_0$. Applying (3), it is found that the fundamental at $2^{19/12} f_0$ is made unresolvable by the 3^{rd} harmonic of f_0. Also, the fundamental at $2^{28/12} f_0$ is made unresolvable by the 3^{rd} harmonic of $2^{10/12} f_0$ and the 5^{th} harmonic of f_0.

Using the type of analysis just outlined, it can be shown that the 57 permissible chords in western polyphonic music fall into two categories [2]:

1. *Completely Identifiable Chords* (43 cases): All fundamentals resolvable, therefore a spectral peak will be found for every F0.
2. *Partially Identifiable Chords* (14 cases): Some fundamentals unresolvable, therefore it is *not* guaranteed that a spectral peak will be found for every F0.

The approach we have developed for chord identification first seeks out all the resolvable fundamental frequencies. It then uses the constraints on allowable combinations of fundamental frequencies in western polyphonic music to seek evidence for the presence of initially unresolvable fundamental frequencies. Our approach is a generalization of the approach presented in [4] for musical chords limited to a maximum of 2 notes.

2. APPROACH

Our approach begins with the observation that the fundamental corresponding to the lowest F0 is always resolvable. This can be seen by noting that if f_0 is the lowest F0, then the rules of western polyphonic music constrain the next highest fundamental to be at $2^{1/12} f_0$ or higher. However, using (3), we note that a component at $2^{1/12} f_0$ cannot render the component at f_0 unresolvable. A second important observation behind our approach is that a fundamental at f_0 can only be made unresolvable by a harmonic of a lower fundamantal. These two observations lead to a straightforward procedure for identifying the spectral peaks corresponding to the resolvable fundamentals.

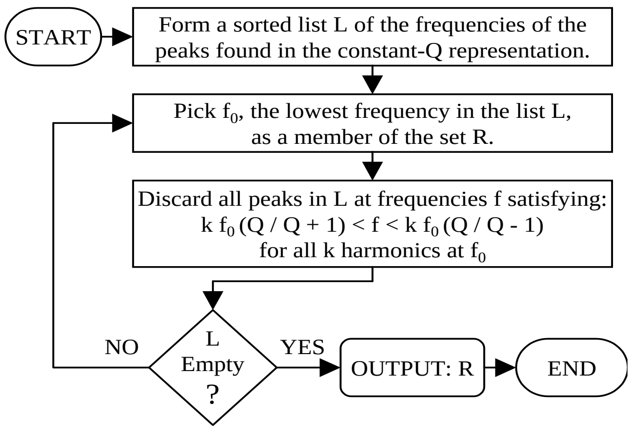

Fig. 1. Strategy for the determination of frequencies of resolvable fundamentals

Resolvable Fundamentals: In seeking the set of resolvable fundamentals, we first form a frequency ordered list L of the frequencies of all the peaks in the CQ spectrum. The lowest frequency in L represents a resolvable fundamental and its frequency is therefore considered an element of a set $\mathcal{R}$ that is ultimately to hold all the F0's measured from the current CQ spectrum. For each harmonic of a measured F0 in $\mathcal{R}$ we discard all frequencies f_1 in L satisfying (3) with f_2 denoting the frequency of the harmonic. If any frequencies remain in our list L, the lowest frequency corresponds to another resolvable fundamental and is therefore appended to the set $\mathcal{R}$. This process is repeated until the list L is empty. The algorithm for the determination of the resolvable fundamentals is illustrated in Figure 1.

Our next step is to identify one or more chords that are a good match to the set $\mathcal{R}$ of measured fundamental frequencies. Given $f0_1$, the lowest measured frequency in $\mathcal{R}$, we may retrieve a set $\mathcal{T}_i$ of theoretical fundamental frequencies corresponding to the i-th chord with lowest fundamental frequency $f0_1$ and $|\mathcal{T}_i| \geq |\mathcal{R}|$ where $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$. Note that a set $\mathcal{T}_i$ of cardinality 3 has the form $\mathcal{T}_i = \{f0_1,\ 2^{\alpha(i,2)} f0_1,\ 2^{\alpha(i,3)} f0_1\}$. We now define a distance measure for comparing $\mathcal{R}$ and each $\mathcal{T}_i$: $d(\mathcal{R},\mathcal{T}_i) = \sum_{j=2}^{|\mathcal{R}|} (f0_j - 2^{\alpha(i,j)} f0_1)^2$. Minimizing $d(\mathcal{R},\mathcal{T}_i)$ with respect to i, we identify the theoretical chords that are the best match in the least-squares sense to the fundamental frequencies in $\mathcal{R}$. The minimization of $d(\mathcal{R},\mathcal{T}_i)$ yields three possible types of results:

1. Only one chord minimizes $d(\mathcal{R},\mathcal{T}_i)$ and $|\mathcal{R}| = |\mathcal{T}_j|$, where $\mathcal{T}_j$ is the set of theoretical fundamental frequencies corresponding to the matching chord.
2. Only one chord minimizes $d(\mathcal{R},\mathcal{T}_i)$ and $|\mathcal{R}| < |\mathcal{T}_j|$, where $\mathcal{T}_j$ is the set of theoretical fundamental frequencies corresponding to the matching chord. In such cases frequency resolution problems cause one of the fundamental frequencies to not appear in the measured set $\mathcal{R}$. However, musical rules dictate that the addtional frequency must be present. See Table 1 for the six cases for which the measured set $\mathcal{R}$ has a smaller cardinality than the matching chord.
3. Two or more chords minimize $d(\mathcal{R},\mathcal{T}_i)$. See Table 2 for the 8 cases in which the measured set $\mathcal{R}$ has multiple chords matching it.

In the first two cases, chord identification is already accomplished. In the third case, the issue of unresolvable fundamentals has to be addressed in order to obtain unambiguous chord identification.

Frequencies in $\mathcal{R}$	**Matching Chords**
f_0, $2^{2/12}f_0$, $2^{9/12}f_0$	f_0, $2^{2/12}f_0$, $2^{9/12}f_0$, $2^{18/12}f_0$
f_0, $2^{4/12}f_0$, $2^{10/12}f_0$	f_0, $2^{4/12}f_0$, $2^{10/12}f_0$, $2^{19/12}f_0$
f_0, $2^{8/12}f_0$, $2^{18/12}f_0$	f_0, $2^{8/12}f_0$, $2^{18/12}f_0$, $2^{27/12}f_0$
f_0, $2^{9/12}f_0$, $2^{14/12}f_0$	f_0, $2^{9/12}f_0$, $2^{14/12}f_0$, $2^{18/12}f_0$
f_0, $2^{9/12}f_0$, $2^{26/12}f_0$	f_0, $2^{9/12}f_0$, $2^{18/12}f_0$, $2^{26/12}f_0$
f_0, $2^{10/12}f_0$, $2^{16/12}f_0$	f_0, $2^{10/12}f_0$, $2^{16/12}f_0$, $2^{19/12}f_0$

Table 1. Six cases in which $|\mathcal{R}| = 3$ but the matching chord has 4 fundamentals

Unresolvable Fundamentals: For the 8 cases in Table 2, we have developed tests which look for evidence for additional F0's that may be present. We start by hypothesizing that the set of resolvable F0's is complete and we look for evidence which may prove the contrary. This evidence may take the form of spectral peaks not attributable to the established F0's but consistent with the presence of additional F0's. Such a possibility exists because an unresolvable component can still give rise to a distinguishable spectral peak in a region consistent with (2). Even in the case when energy from two unresolvable components is merged to form a single peak, we have the option of computing the spectrum with a higher Q in an attempt to avoid the merging of the energy components in the spectral domain. However, care has to be exercised because of the longer impulse responses corresponding to higher values of Q. Based on this general approach and on our empirical experimentation, we have developed strategies to identify the additional F0's for each of the 8 cases in Table 2.

The general procedure for carrying out the test cases involves a search for evidence supporting the existence of an additional fundamental at a frequency f_a. We first identify two frequencies f_1 and f_2 which represent harmonics of already-identified fundamentals such that f_1 is the nearest harmonic frequency smaller than f_a and f_2 is the nearest harmonic frequency greater than f_a. Next, we locate in the CQ transform the peak closest to f_1 and the peak closest to f_2. Within the region between these two peaks we look for evidence of another peak corresponding to an additional fundamental at frequency f_a. If a peak is found, we conclude that the fundamental at f_a is present. Otherwise, in certain cases, it is feasible to recompute the spectrum with a $Q = Q_{new}$ which is sufficient to separate any potentially merged peaks. Using the recomputed spectrum, we repeat the test, where now, if a peak is not found, we conclude that f_a is actually not present. We justify this procedure by noting that even though the peak at f_a corresponds to a theoretically unresolvable component, there is still a likelihood for there to be a peak corresponding to f_a. The likelihood is greater when the CQ transform is recomputed with Q_{new} (a higher Q).

In Table 3 we specify the parameters needed to carry out the

	Frequencies in $\mathcal{R}$	Matching Chords
1	f_0	A. f_0 B. f_0,$2^{1/12}f_0$ C. f_0,$2^{11/12}f_0$
2	f_0,$2^{6/12}f_0$,$2^{15/12}f_0$	A. f_0,$2^{6/12}f_0$,$2^{15/12}f_0$ B. f_0,$2^{6/12}f_0$,$2^{15/12}f_0$,$2^{20/12}f_0$
3	f_0,$2^{9/12}f_0$,$2^{17/12}f_0$	A. f_0,$2^{9/12}f_0$,$2^{17/12}f_0$ B. f_0,$2^{9/12}f_0$,$2^{17/12}f_0$,$2^{27/12}f_0$
4	f_0,$2^{10/12}f_0$	A. f_0,$2^{10/12}f_0$ B. f_0,$2^{10/12}f_0$,$2^{19/12}f_0$,$2^{28/12}f_0$
5	f_0,$2^{9/12}f_0$	A. f_0,$2^{9/12}f_0$ B. f_0,$2^{9/12}f_0$,$2^{18/12}f_0$
6	f_0,$2^{8/12}f_0$	A. f_0,$2^{8/12}f_0$ B. f_0,$2^{8/12}f_0$,$2^{18/12}f_0$,$2^{27/12}f_0$
7	f_0,$2^{3/12}f_0$,$2^{8/12}f_0$	A. f_0,$2^{3/12}f_0$,$2^{8/12}f_0$ B. f_0,$2^{3/12}f_0$,$2^{8/12}f_0$,$2^{18/12}f_0$
8	f_0,$2^{8/12}f_0$,$2^{15/12}f_0$	A. f_0,$2^{8/12}f_0$,$2^{15/12}f_0$ B. f_0,$2^{8/12}f_0$,$2^{15/12}f_0$,$2^{18/12}f_0$

Table 2. Completing the initial list of F0's - Two possible completions

tests for all but one of the eight cases enumerated in Table 2. In case 1, the peak at f_0 corresponds to a resolvable fundamental in this 2-note chord, and therefore if $2^{1/12} f_0$ is present, the spectral energy due to f_0 will not merge with the energy at $2^{1/12} f_0$ (i.e there is no need to use a higher Q in this case). For case 4 in Table 3 the likelihood of finding peaks corresponding to the additional F0's between harmonics of the resolvable fundamentals is very low. Instead, the procedure for case 4 involves searching the CQ spectrum to see if in addition to the spectral peak corresponding to $3\times 2^{10/12}f_0$, there is a nearby peak corresponding to $2^{28/12}f_0$. If a peak is found, we conclude that $2^{28/12}f_0$, the additional F0, is present. If a peak is not found, we conclude that $2^{28/12}f_0$ is not present. The justification for this procedure is that the peak at $3\times 2^{10/12}f_0$ corresponds to a resolvable component unless the peak corresponding to the component at $2^{28/12}f_0$ is present. The spectral energies of the two peaks will not merge in this case, since although the component at $3\times 2^{10/12}f_0$ is affected by the component at $2^{28/12}f_0$, the reverse is not true.

3. IMPLEMENTATION ISSUES

Gaussian Constant-Q Filtering: The analysis in the constant-Q filterbank is carried out with Gaussian filters which were chosen on the basis of the well-known property that they have the least time-frequency uncertainty [5]. The center frequencies of the filters in the filterbank are uniformly spaced along the frequency axis and the impulse response of the i-th filter is given by:

$$h_i[n] = \begin{cases} A\exp\{-\alpha n^2\}\exp\left\{j\frac{2\pi f_i}{F_s}n\right\}, & |n| \leq n_i \\ 0, & |n| > n_i \end{cases}.$$

Here f_i is the center frequency of the filter, F_s is the sampling rate, n_i is the time index before which the magnitude of the impulse response decays to a small value ϵ, A is a scaling factor which ensures that the filter has unit energy, and α is a parameter that controls the bandwidth of the filter. To ensure that the filter has a bandwidth which follows a constant-Q rule, we utilized the filter's frequency response [5] to arrive at the following relation:

$$\alpha = \frac{1}{2\ln(\epsilon)}\left[\frac{\pi f_i}{QF_s}\right]^2.$$

Peak Picking Algorithm: An initial set of peaks is established by detecting local maxima in the spectrum. Any peaks that are 40dB below the highest peak are eliminated from this set. Starting with the highest energy peak as the "reference peak," we proceed to eliminate all peaks that are smaller than the magnitude of the filterbank's frequency response (a function of the center frequencies of the filters in the filterbank) to a complex exponential with an amplitude and frequency equal to that of the reference peak. The rationale for this procedure is that peaks smaller than the frequency response magnitude are essentially sidelobes corresponding to the reference peak. We keep repeating this procedure, using as the reference peak the peak whose energy is largest from amongst the un-eliminated peaks that have not already been used as reference peaks. Once all un-eliminated peaks have been used as reference peaks, we utilize one final procedure to eliminate each peak that is theoretically unresolvable from another peak of 12dB greater energy.

Bin Resolution: This is the issue of how far apart should we have the center frequencies of consecutive filters in the filterbank. We note that in accordance with (3), a component at frequency f_2 is guaranteed not to be affected by a component at frequency f_1 if the distance between f_1 and f_2 is greater than

$$D_{min} = Max((1 - \frac{Q}{Q+1})f_1, (\frac{Q}{Q-1} - 1)f_1).$$

To be able to detect spectral peaks corresponding to such components we must have a frequency spacing between the center frequencies of consecutive filters of at least $\frac{D_{min}}{2}$. Assuming that the smallest possible value for f_1 is 196Hz (violin G3) and using a Q of 17, the corresponding value of $\frac{D_{min}}{2}$ is approximately 5Hz. For higher values of f_1, $\frac{D_{min}}{2}$ would be larger. In our implementation the filters' center frequencies were kept uniformly spaced and so we used a 5Hz spacing throughout the filterbank.

Case #	$\mathbf{f_1}$	$\mathbf{f_2}$	$\mathbf{f_a}$	$\mathbf{Q_{new}}$
case 1	f_0	$2f_0$	$2^{1/12}f_0$	–
case 2	$3f_0$	$4f_0$	$2^{20/12}f_0$	38
case 3	$4f_0$	$5f_0$	$2^{27/12}f_0$	50
case 5	$2f_0$	$3f_0$	$2^{18/12}f_0$	38
case 6	$2f_0$	$3f_0$	$2^{18/12}f_0$	38
case 7	$2\times 2^{8/12}f_0$	$3f_0$	$2^{18/12}f_0$	38
case 8	$2^{15/12}f_0$	$3f_0$	$2^{18/12}f_0$	38

Table 3. Parameter specification for test cases

4. VALIDATION EXPERIMENTS

A software implementation of our approach was used to successfully validate its performance on a representative data set. We generated 2,266 signals corresponding to all valid 2-note, 3-note

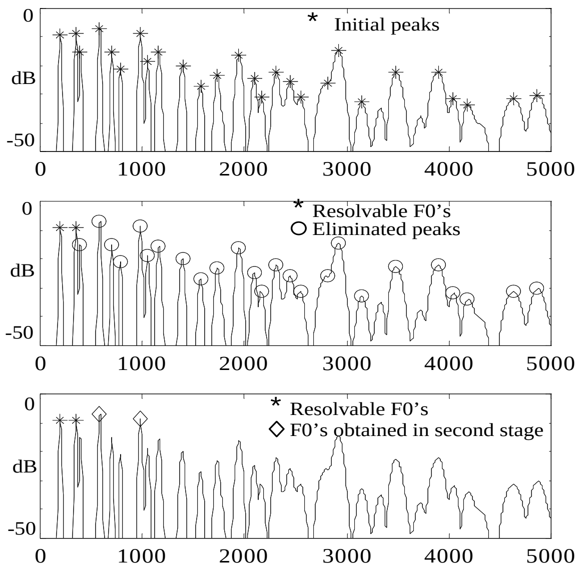

Fig. 2. The constant Q spectrum of a musical signal and the peaks identified during various stages of processing

and 4-note chords that a violin quartet can play. This included all the chords that can be composed from notes whose fundamental frequencies range from 196Hz(G3) to 3,951Hz(B8), generally accepted limits for violins. Each test signal was 0.5 seconds long and sampled at $F_s = 20480$Hz. The relative energies of the various harmonics of each note were determined in accordance with well- established spectral models for violin sounds [6] which suggest a +6dB/octave decrease in spectral energy at frequencies up to 3KHz and 15dB/octave thereafter. The net energy in each note was normalized to a fixed value in order to simulate conditions of "comparable loudness" for each of the notes in a chord. Any harmonics at frequencies above $\frac{F_s}{2}$ were eliminated (even though they were included in the energy normalization process). We did not include in our test signals instances of the half-tone case (the closest possible fundamentals) in situations where one of the F0's is below 200Hz. This is because filters with Q=17 and center frequency below 700Hz have impractically long impulse responses for the music application. Therefore, in our implementation of the CQ spectrum, filters centered at frequencies below 700Hz have the same bandwidth as that of a filter centered at 700Hz. However, for F0's below 200Hz, this higher BW results in low resolution and the inability to resolve an F0 and its half-tone with a reasonable Q factor.

In Figure 2 we present an example of the type of results obtained during the validation process. The musical signal in this case is a combination of four notes with $f_0 = 196.0$Hz ($G3$), and the remaining three notes: 349.2Hz ($F4$), 587.3Hz ($D5$) (vibrato rate of 10Hz) and 987.8Hz ($B6$) (vibrato rate of 10Hz). Note that this example corresponds to case 4B listed in Table 2. The top plot in Figure 2 shows the CQ spectrum with the initial set of peaks identified by our peak-picking algorithm. The second plot for the same CQ spectrum shows the peaks eliminated through the procedure of Figure 1 as well as the two peaks identified to correspond to the fundamentals at 196Hz and 349.2Hz. The final plot also shows the additional two peaks identified during the second phase of our approach.

5. ACKNOWLEDGEMENTS

We would like to thank S. Hebsur, R. Mani, and G. Taibi for their assistance during the course of the reasearch reported in this paper.

6. REFERENCES

[1] P. Goetschius. *The Theory and Practice of Tone Relations*. Greenwood Press, Conn., 1972.

[2] R. Mani. *Time-Frequency Signal Representation for Polyphonic Music*. PhD thesis, Boston University, Boston, May 1998.

[3] J. Brown. Calculation of a constant Q spectral transform. *Journal of the Acoustical Society of America*, 89(1):425–434, 1990.

[4] R. Mani and S. H. Nawab. Knowledge-based processing of multicomponent signals in a musical application. *Signal Processing – Elsevier*, 74(1999):47–69, 1999.

[5] F. J. Harris. On the use of windows for harmonic analysis with the discrete-fourier transform. In *Proceedings of the IEEE*, volume 66, pages 51–83. IEEE, 1995.

[6] N. Fletcher and T. Rossing. *The Physics of Musical Instruments*. Springer, New York, New York, 1998.