Automatic Interval Naming Using Relative Pitch *

David Gerhard
School of Computing Science
Simon Fraser University
Burnaby, BC V5A 1S6
E-mail: dgb@cs.sfu.ca

Abstract

Relative pitch perception is the identification of the relationship between two successive pitches without identifying the pitches themselves. Absolute pitch perception is the identification of the pitch of a single note without relating it to another note. To date, most pitch algorithms have concentrated on detecting the absolute pitch of a signal. This paper presents an approach for relative pitch detection, and applies this approach to the problem of detecting the musical interval between two acoustic events. The approach is presented as it applies to the western system of music.

1. Introduction

The human auditory system allows humans to perceive differences in air pressure and attach meaning to different patterns—we hear sounds. Everything that humans hear is an interpretation of the time-varying air pressure on the ear drum. Consequently, concepts like pitch and timbre are interpretations made somewhere between the eardrum and the conscious mind. Interpretations such as these sometimes do not fully reflect the real world—in human vision, for example, metamerization occurs when two objects with different surfaces are perceived as the same colour, and colour constancy occurs when two objects with the same surface are perceived as different colours [6] [1]. In the same way, two sounds of the same frequency can "appear" to have different pitches, depending on other qualities of the sound such as loudness and timbre [9]. Audio illusions occur when an ambiguous audio stimulus is resolved by the brain [2], just as optical illusions can make humans perceive three dimensions in a two-dimensional surface.

Absolutely Relative

Pitch is an important part of understanding and perceiving western music. Much work has been done recently on automatic music transcription, where musical audio is translated directly to a score representation. Most researchers approach this problem by approximating the fundamental frequency (f_0) of the sound at each point and using that to approximate the absolute pitch of the

music at that point. One problems with this approach is the subjectivity of pitch. Another problem relates to the fact that most music consists of many notes being played at the same time, called polyphonicity.

Relatively Standard?

Absolute pitch is a subjective quality. From 1739 to 1879, the standard frequency for the A above middle C, cited from piano and organ manufacturers, varied from 392 Hz to 563 Hz, or from G below today's standard A to slightly above the C♯ above today's standard A [5] (most manufacturers today use a 440 Hz A as standard). If an instrumental combo has a large instrument like a piano or organ, then the other instruments will tune to it, resulting in the entire combo playing in that tuning.

What has not changed as much in the centuries of western music is the intervals between standard pitches, or *relative* pitch. The pythagorian scale and the just scale date from antiquity and relate tones using the ratio of their frequency. Newer scales such as the meantone scale and the scale of equal temperament are attempts to make the earlier scales playable in any key. The pitch intervals in these new scales are very similar to the old scales, but modern string orchestras sometimes play leading pitches higher, to more accurately approximate the older scales.

When a person hums a tune in their head or out loud, they are using relative pitch. It doesn't matter what pitch the person uses to begin the song, the tune is recognizable as long as the intervals between the notes are reproduced accurately. The first scale that many children learn is the "do-re-mi-fa-so-la-ti-do" scale. The base note "do" can vary widely, but the relationships *between* notes are well defined and easily learned. Most people can sing a "do-re" or a "do-so" for example. Absolute pitch can be learned, but it is much more difficult than learning relative pitch, and, once learned, this absolute pitch recognition is much slower and less accurate than inborn absolute pitch recognition [8].

Many and One

Automatic music transcription comes in two flavors: polyphonic and monophonic. Polyphonic music transcription is the problem of writing down the score of a piece of music when more than one instrument is playing. It is the more common problem—in western music, there are usually many instruments playing at the same time. It is also a more difficult problem, without a complete solution to date.

In contrast, monophonic music transcription is relatively simple. If there is only one instrument playing, it is a matter of finding the pitch of the instrument at all points, finding where the notes change, and what the time signature and key signature of the piece is. Some of these problems are harder than others, but a complete system for monophonic music transcription was presented in 1986 [10].

Compute me a Tune

When working on transcription systems, whether polyphonic or monophonic, most researchers start with absolute pitch detection, and work from there. Automatic absolute pitch detection is a very difficult problem, even for a monophonic signal. Research into automatic absolute pitch detection has lead to many different methods, each with their related difficulties [4] [7] [11]. If the tone is

pure, without any harmonics and without noise, then the computer can approximate the frequency of the tone by counting how many beats occur in a second, and approximate the pitch from that using whatever subjective standard is in style today.

It is seldom that easy. Most western instruments create very complex tones, with many harmonics and overtones, and there is usually noise present in the signal. Researchers have taken to using spectral transforms, which measure how much of each frequency there is in the signal, and then approximating the fundamental frequency by looking for the lowest frequency component that is stronger than a given threshold, or by looking for peaks in the spectrogram. These transforms are based on a specific frequency, so the results are related to that base frequency, and many difficult calculations must be done to extract an *approximation* of the frequency of the signal.

In contrast, the spectrogram transforms are well suited to discovering the *relative* pitch of a signal. The base frequency that these transforms use does not hinder the calculation of the pitch interval because both notes use the same transform with the same base frequency, and it is factored out of the calculation.

The World Over

This paper is limited to the study of western music, which is based on particular scales and rhythms. Western music is clearly not a complete model for all human music, as most cultures have their own musical systems based on different scales and rhythmic patterns, some entirely rhythmic and some entirely tonal. The concepts presented in this paper could be extended to apply to other cultural musical systems.

Music *perception* is culturally based the same way that music production is culturally based. The music that people hear as they grow and develop becomes the reference point for the music they find appealing in maturity. For this reason, any study of music should be qualified by indicating the musical system being studied.

2. Relative Pitch Perception

Some Notation

There are different methods used to write information about pitches and their composition. For reference, here is the notation used in this paper. Many of the concepts, such as the harmonic series, are described later.

X_n A note in the nth standard piano octave. For example, C_4 is C in the 4th octave, or middle C.

S_n The nth note in the scale S. For western music, in the scale of equal temperament, there are 12 semitones in a scale, so $n = 0, 1, \ldots, 11$. S_{12} is an octave above S_0, the *tonic*, or root note.

f_0 The fundamental frequency of an audio signal.

$f_0(X_n)$ The fundamental frequency of a note in the nth octave. For example, according to modern tuning, $f_0(A_4) = 440$ Hz.

$h_k(X_n)$ The frequency of the kth harmonic of a note X_n.

$a_k(X_n)$ The amplitude of the kth harmonic of a note X_n.

$\langle a_1, a_2, a_3, \ldots \rangle$ A harmonic series, or spectrum of amplitudes of harmonics of a note.

$Q \nearrow R$ The named interval between two notes Q and R, such as "semitone","tone" or "major third".

Logarithmic Perception of Pitch

Humans perceive pitch on an approximately logarithmic frequency scale. If $f_0(A_4) = 440$ Hz, then $f_0(A_5) = 880$ Hz, and $f_0(A_3) = 220$ Hz. An octave increase in the pitch of a signal corresponds to about a doubling of the f_0 of that signal. This relationship is slightly distorted at the high end of the frequency scale as well as the high end of the loudness scale, but in the mid-range of human hearing, this logarithmic correspondence holds [5].

At lower frequencies, a semitone corresponds to a smaller pitch jump than at higher frequencies. For example, $f_0(F_2) - f_0(E_2) = 87.31 - 82.41 = 4.90$ Hz difference, and $f_0(F_6) - f_0(E_6) = 1396.91 - 1318.51 = 78.40$ Hz difference, using $f_0(A_4) = 440$ Hz.

Linearity of Harmonics Within a Pitch

When an instrument is played, it sets up vibrations in the air at f_0 of the note being played, as well as vibrations at $2f_0$, $3f_0$ etc. These higher frequency vibrations are called *harmonics*, and they are what makes a trumpet sound different than an organ. These harmonics are equally spaced in the frequency domain, and can be collectively referred to as a *harmonic series*. The first harmonic is at the same frequency as the fundamental, so for any note, $f_0 = h_1$. The locations of all harmonics of a note can be generated from f_0 using

$$h_k = k \cdot f_0 \tag{1}$$

Harmonic Series. The harmonics of a note also have associated amplitudes, corresponding to how much of each harmonic is present in the note. If an instrument generates the fundamental frequency only, with no harmonics, then a_1, the amplitude of h_1, would be the amplitude of the signal, and the amplitudes of the other harmonics, $a_k(k \geq 2)$ would be zero. This is an example of the *spectrum* of an instrument, which is the sequence of amplitudes of the harmonics that the instrument generates. A typical spectrum is shown in Fig. 1. The spectrum of an instrument is related to the *timbre*, or characteristic sound quality, of that instrument. Instruments have different sounds because they have different spectra. For examples of the spectra of different instruments, including spectra of the human voice, see [9].

The spectrum for a particular instrument also depends on the note being played on that instrument. The general shape of the spectrum might be the same for all notes from the same instrument, but the values of the coefficients and their locations will be different.

Dropped Harmonics. Not all musical signals have all harmonics present. A sinusoidal signal has $\langle a_1, 0, 0, 0, \ldots \rangle$, with a_1 the amplitude of the sinusoid. A square wave has $\langle a_1, 0, a_3, 0, a_5, 0, \ldots \rangle$,
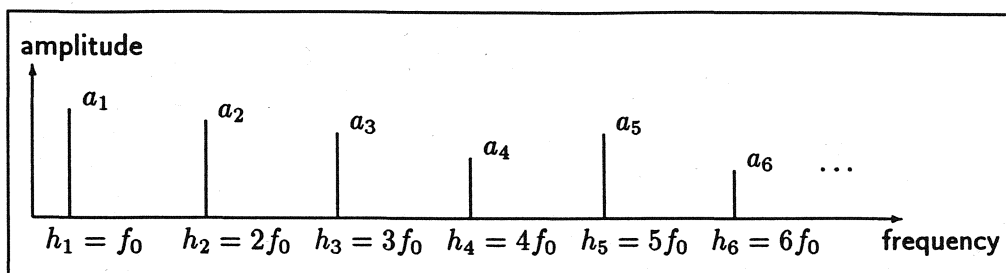
Figure 1: Typical spectrum of a note with fundamental frequency f_0.

where $a_{2k} = 0$. This phenomenon, where specific harmonics have zero amplitude, is called "dropped harmonics". Many artificial and computer-generated signals have dropped harmonics, but few naturally occurring signals do, with the exception of the above mentioned sinusoid.

Convergence. The amplitude of every harmonic in a series is non-negative ($a_i \geq 0$), and every harmonic series is convergent to zero ($\lim_{i \to \infty} a_i = 0$), but is not necessarily monotonic ($a_i \geq a_{i+1}$). The harmonics of a note that can be detected above the ambient noise in a signal depends on the amplitude of the harmonics, the level of ambient noise, and the pitch of the note. If the pitch is very high, only the first few harmonics will be detectable in the spectrogram, because as the pitch increases, the distance between the harmonics increases as well. The harmonics of a note at 880 Hz will be twice as far apart as the harmonics of a note at 440 Hz.

There are advantages and disadvantages of natural signals for the approach to interval detection presented in this paper. Natural signals tend to have more noise, making only the first few harmonics detectable in a spectrogram, depending on the pitch. Conversely, very few natural signals have dropped harmonics.

3. The Approach

This approach to musical interval detection takes advantage of the fact that while notes on a musical scale are perceived on an approximately logarithmic scale, the harmonics of a single note are approximately linearly related. This means that when two notes are played, some harmonics will overlap at specific points in the frequency domain. Which harmonics overlap will indicate the interval between the notes being played.

Two Scales

The scale of *equal temperament* is the musical scale in common usage in western music today, and it replaces the more accurate but less adaptable scale of *just intonation*. The f_0 of each note in the equal scale is calculated exponentially from f_0 of the tonic, using Equation 2. Recall that $f_0(S_n)$ is the fundamental frequency of the nth note in a scale, and $f_0(S_n)$ is the fundamental frequency of the tonic, or starting note. For the equal scale,

$$f_0(S_n) = f_0(S_0) \times 2^{\frac{n}{12}}. \tag{2}$$

Just Interval	Just Ratio	Closest Equal Ratio	Equal Interval
Unison	$1:1 = 1.0$	$1.0 = 2^{\frac{0}{12}}$	Unison
Semitone	$16:15 = 1.0666\bar{6}$	$1.05946 = 2^{\frac{1}{12}}$	Semitone
Minor tone	$10:9 = 1.1111\bar{1}$	$1.12246 = 2^{\frac{2}{12}}$	Whole tone
Major tone	$9:8 = 1.125$	$''$	$''$
Minor 3rd	$6:5 = 1.2$	$1.18921 = 2^{\frac{3}{12}}$	Minor 3rd
Major 3rd	$5:4 = 1.25$	$1.25992 = 2^{\frac{4}{12}}$	Major 3rd
Perfect 4th	$4:3 = 1.3333\bar{3}$	$1.33484 = 2^{\frac{5}{12}}$	Perfect 4th
Augmented 4th	$45:32 = 1.40625$	$1.41421 = 2^{\frac{6}{12}}$	Augmented 4th, or
Diminished 5th	$64:45 = 1.4222\bar{2}$	$''$	Diminished 5th
Perfect 5th	$3:2 = 1.5$	$1.49831 = 2^{\frac{7}{12}}$	Perfect 5th
Minor 6th	$8:5 = 1.6$	$1.58740 = 2^{\frac{8}{12}}$	Minor 6th
Major 6th	$5:3 = 1.6666\bar{6}$	$1.68179 = 2^{\frac{9}{12}}$	Major 6th
Harmonic Minor 7th	$7:4 = 1.75$	$1.78179 = 2^{\frac{10}{12}}$	Minor 7th
Grave Minor 7th	$16:9 = 1.7777\bar{7}$	$''$	$''$
Minor 7th	$9:5 = 1.8$	$''$	$''$
Major 7th	$15:8 = 1.875$	$1.88775 = 2^{\frac{11}{12}}$	Major 7th
Octave	$2:1 = 2.0$	$2.0 = 2^{\frac{12}{12}}$	Octave

Table 1: Fundamental Frequency Ratios in the Scales of Just Intonation and Equal Temperament.

In particular,

$$
\begin{aligned}
f_0(S_{12}) &= f_0(S_0) \times 2^{\frac{12}{12}} \\
&= 2f_0(S_0),
\end{aligned}
\tag{3}
$$

which shows that the octave tone is twice the frequency of the tonic, as expected.

The scale of just intonation is a perfect ratio scale, with f_0 of every note in the scale a whole number ratio from $f_0(S_0)$. The problem with the just scale is that the notes are only valid for a specific key signature, and instruments need to be adjusted when played in a different key.

The equal scale allows instruments to be played in all keys without re-tuning. It is a compromise from the scale of just intonation, and as a result, all of the notes are slightly out of tune. The western ear has become accustomed to equal temperament, and the tuning differences are hardly noticeable.

The intervals in the just scale are presented in Table 1, along with their numerical ratios. For each interval in the just scale, the closest numerical ratio and corresponding interval in the scale of equal temperament are also presented.

Depending on the role of a note in the scale, it can have one of several f_0's in the just scale, which is why the just scale is only valid for one key. As an example, the note "E" occurs in both the key of C major and the key of G major. If $f_0(C_4) = 261.63$ Hz, then in the just scale, $f_0(E_4)$, being

the major third, will be $\frac{5}{4} \cdot 261.63 = 327.04$ Hz. In the key of G major, however, with a tonic of $f_0(G_3) = 196.00$ Hz, $f_0(E_4)$ is a major sixth and will be $\frac{5}{3} \cdot 196.00 = 326.67$ Hz. The difference between these two frequencies is .37 Hz, which doesn't seem like much, but if these two notes were to be played together, an undesirable interference pattern would occur.

In the equal scale, $f_0(E_4)$ calculated from $f_0(C_4)$ is $261.63 \times 2^{\frac{4}{12}} = 329.63$, and calculated from $f_0(G_3)$ is $196.00 \times 2^{\frac{9}{12}} = 329.63$, the same value, slightly higher than both E's in the just scale. Thus, the intervals are slightly out of tune from the just scale, but the notes are in tune with each other, allowing musicians to change keys between pieces or in the middle of a piece without re-tuning their instruments.

The Technique

This approach uses two facts about the harmonics of a pair of notes to determine the interval between the notes. These facts are treated independently in the two following methods for relative pitch approximation, and the results of one can be used to confirm the results of the other.

For any two notes Q, R with fundamental frequencies $f_0(Q) \leq f_0(R) \leq 2f_0(Q)$ (R higher than Q but within an octave),

Method 1 *Normalize the spectrum of harmonics of the notes Q and R such that $h_1^Q = 1$ and $h_2^Q = 2$. Then if $1 \leq h_1^R \leq 2$, h_1^R is the ratio between the fundamental frequencies of the notes, $\frac{f_0^R}{f_0^Q}$, and can be used to approximate the equal temperament interval of the note pair, from Table 1.*

Notes. Normalization of the harmonic series corresponds to dividing the frequency of each harmonic by the fundamental frequency, so that $h_n^Q = n$. If the exact frequencies of the harmonics are not known, as is often the case when trying to approximate the pitch, the whole numbers can be assigned directly to the spectrogram output. When the normalized frequency axis for a note is used to read the location of a *different* note, the result is to read the frequency ratio between the notes, which corresponds directly to the interval between the notes.

Method 2 *Find two harmonics, h_i^Q and h_j^R, one from each note, which occur at the same frequency $h_i^Q = h_j^R$. Then the ratio $i : j$ can be used with Table 1 to approximate the just intonation interval of the note pair.*

Notes. When particular harmonics of two different notes occur at the same frequency, the ratio between the fundamental frequencies of these notes is directly related to the ordinals of the overlapping harmonics. $h_i^Q = h_j^R$ implies that $i \cdot f_0^Q = j \cdot f_0^R$, from Eq. 1, which further implies that $\frac{i}{j} = \frac{f_0^R}{f_0^Q}$. This means that the ratio of the ordinals of the overlapping harmonics gives the frequency ratio between the notes, which corresponds directly to the interval between the notes.

Fig. 2 shows the use of Method 1. Here, the frequency axis is normalized such that the first two harmonics of Q occur at 1 and 2. The first harmonic, or fundamental, of R is seen to fall at 1.25,
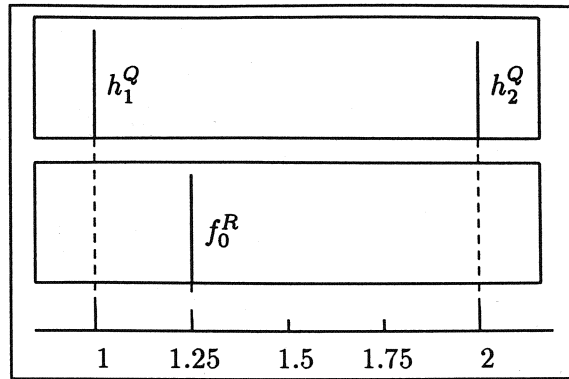
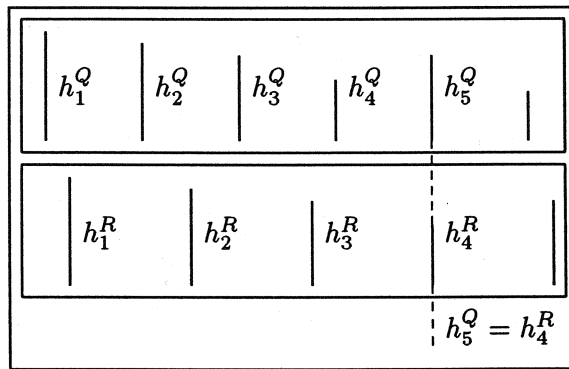Figure 2: Comparison of f_0^R to h_1^Q and h_2^Q, indicating that $Q \nearrow R$ is a major third.

Figure 3: Matching h_5^Q to h_4^R, indicating that $Q \nearrow R$ is a major third.

or a quarter of the way from h_1^Q to h_2^Q. Combined with Table 1, this is sufficient information to deduce that $Q \nearrow R$ is a major third.

Fig. 3 shows the use of Method 2. In this case, the first 6 harmonics of Q are detectable, as are the first 5 harmonics of R. The 5th harmonic of Q occurs at the same location on the frequency axis as the 4th harmonic of R, and, combined with Table 1, this is sufficient information to deduce that $Q \nearrow R$ is a major third.

Compounding Intervals

These proposed methods are not specifically designed to handle the case where the frequency of the first harmonic of R is greater than the frequency of the octave above Q, i.e. if $h_1(R) > 2h_1(Q)$. It is necessary to augment the methods to handle this case, but the required modifications are minimal.

Augmenting Method 1. The normalization used in the first method applies to the entire range of frequencies, and is not restricted to the interval between $h_1(Q)$ and $h_2(Q)$. The frequency ratio will still be valid for larger intervals, but the naming of these intervals is not handled by Method 1. The modification is to name the interval as a number of octaves plus an interval from Table 1. If the ratio can be written or approximated in the form $2^{\frac{k+12n}{12}}$ then the interval is n octaves, plus the interval in Table 1 corresponding to the ratio $2^{\frac{k}{12}}$.

The augmentation can be demonstrated in an example: If $h_1(R) = 2.6697$ on the normalized scale of Q, this is best approximated by the exponential $2^{\frac{17}{12}}$, which is the same as $2^{\frac{(5+12)}{12}}$, therefore the interval is identified as a perfect fourth plus an octave.

This augmentation also allows Method 1 to detect intervals *less* than an octave. If $h_1(R)$ falls *below* $h_1(Q)$, Method 1 is still valid and the interval can be considered to be an octave *less* than the interval found in Table 1. For example, If $h_1(R) = 0.5$ on the normalized scale of Q, this is best approximated by the exponential $2^{-1} = 2^{\frac{(0-12)}{12}}$, therefore the interval is identified as unison minus an octave.

Augmenting Method 2. For Method 2 to be able to identify intervals above the octave, Table 1 must be extended to contain all these extra ratios. Since an increase of an octave corresponds to about a doubling of f_0, doubling each frequency ratio in the table corresponds to increasing each frequency ratio by an octave: if 5:4 corresponds to a major third, then 10:4 corresponds to an octave plus a major third.

Method 2 can then identify intervals larger than the octave by finding coincident harmonics and comparing the ordinals to those in Table 1, as well as whole number multiples of the intervals in Table 1. It is impossible to check every whole number multiple of every interval, so a limit should be imposed to make the method computationally tractable. This is not unreasonable, considering that the average human ear can only detect frequencies below about 20,000 Hz.

As with Method 1, this augmentation can allow Method 2 to detect intervals below unison. If 4:3 corresponds to a perfect fourth, then 2:3 corresponds to a perfect fourth minus an octave. It is impossible to detect a coincidence between the 4th harmonic of one note and the 2.5th harmonic of another note, as would be required to detect a major third minus an octave, and this limits the usability of Method 2 on intervals less than unison.

Another way to handle intervals less that unison is to reverse the order of the notes. If using Q as the root note yields a ratio less than 1, use R as the root note instead, and employ Method 2 as usual. This provides the interval $R \nearrow Q$, and $Q \nearrow R$, if needed, can be obtained by inverting the detected ratio.

With these augmentations, the proposed methods can handle any interval. The restriction placed on the methods that $f_0(Q) \leq f_0(R) \leq 2f_0(Q)$ can be lifted.

4. Discussion

These are independent methods of using relative analysis of the harmonics to determine the pitch interval. If the two methods yield consistent results, there is reasonable confidence that the interval identification is accurate. An inconsistency in the result might indicate that one or the other of the auditory events did not have a pitch or that there were dropped harmonics or some other error. In that case, further analysis such as noise filtering methods or a different spectral transform could be performed. It is important to identify which harmonics are present and detectable before applying either of the methods.

The proposed approach could be used for absolute pitch recognition, by assuming a beginning note

and identifying each successive note from the intervals between it and the note before it. The pitches thus identified can be compared to the original pitches and corrected up or down to provide a best fit of the melody for the length of the piece.

Overcoming Some Limitations

Polyphonicity. Audio signals with more than one note playing at the same time are difficult to analyze in terms of harmonic series. When more than one harmonic series exists in the spectrogram, it is not clear which harmonics belong to which series, until some analysis is done. Further research may produce an algorithm that is capable of separating a chord into component notes, and such an algorithm might be based on finding and subtracting harmonic series in the audio signal. If a harmonic series is detected, from the regularly-spaced spikes in the frequency domain, it can be filtered out and identified as a note. Then the remainder of the signal can be treated the same way until there are no more spikes in the signal.

Inaccuracy of the Spectrogram. Another problem for this approach is that harmonic components in a spectrogram representation rarely occur at a single isolated frequency. They usually are manifest as distributions around a central frequency. For this reason they are difficult to localize, and there is often error between the detected and actual location of any harmonic. The locations of the harmonics are known to be more or less a linear progression, so a linear best fit could be done on the estimated locations of the harmonics, increasing the accuracy of the approximation.

Undetectable Harmonics. Most natural musical signals contain harmonics with amplitude smaller than the amplitude of the ambient noise in the signal. Such harmonics are undetectable by present spectrographic techniques. If there are harmonics that are not detectable, but are needed for one of the methods to work properly, they can be approximated using the existing harmonics of the note and Eq. 1. A linear best fit can be performed on the detected harmonics, and the location of undetected harmonics can be extrapolated from this linear best fit model. As an example, if the first 3 harmonics of a note are present, an approximation of h_4 could be made using the average of the two differences $h_2 - h_1$ and $h_3 - h_2$ to approximate the difference $h_4 - h_3$. This difference would be added to h_3 to provide an approximation for h_4, and similarly for h_5, h_6 and so on. More detectable harmonics will increase the accuracy of the approximation of undetected harmonics.

Non-overlapping Harmonics. Most modern instruments play in the scale of equal temperament, where $\frac{f_0^R}{f_0^Q}$ is not necessarily an exact whole number ratio. In this case, the whole number ratio that is the closest to the measured ratio will be taken as the ratio for the interval. This will work well for Method 1, but it could be problematic for Method 2, where harmonics are not likely to be exactly coincident. Finding the pair of harmonics that are closest together is not trivial, especially if not all harmonics are present in the measured spectrogram. This is a case where the consistency between the methods is particularly useful. The relationship between the the two *sequences* of harmonics could also be used in this case. If no two harmonics are coincident, then the difference between pairs of harmonics could be measured and analyzed: if h_7^R is fairly close to h_4^Q, and h_8^R is very close to h_5^Q, but h_9^R is a little further away from h_6^Q again, it is probable that the ratio in question is a minor sixth, with ratio 8:5.

5. Conclusion

An approach for pitch interval detection is presented, on the premise that the human auditory perceptual system is better at relative pitch detection than absolute pitch detection, which suggests that the task of interval detection might be easier than the task of absolute pitch detection. Two methods are used to approximate the ratio between the fundamental frequencies of two temporally separated notes. Method 1 compares the location of the fundamental frequency of the second note with the locations of the first two harmonics of the first note, indicating an interval in the scale of equal temperament. Method 2 identifies harmonics of the two notes that are coincident, indicating an interval on the scale of just intonation.

References

[1] Brainard, David H. and Wandell, Brian A.. Analysis of the retinex theory of color vision. Journal of the Optical Society of America A, Vol. 3 No. 10, pp1651-1661, 1986.

[2] Bregman, Albert S. *Auditory Scene Analysis* Cambridge: MIT Press, 1990.

[3] Cooper, William E. and Sorenson, John M. *Fundamental Frequency in Sentence Production.* New York: Springer-Verlag, 1981.

[4] Dorken, E. and Nawab, S. H.. Improved musical pitch tracking using principal decomposition analysis. IEEE-ICASSP 1994.

[5] Eargle, John M. *Music, Sound and Technology.* Toronto: Van Nostrand Reinhold, 1995.

[6] Hubel, David H. and Wiesel, Torsten N.. Brain Mechanisms of Vision. Scientific American, Vol. 241 No. 3, pp150-162, 1979.

[7] Katayose, Haruhiro. Automatic Music Transcription. Denshi Joho Tsushin Gakkai Shi, Vol. 79, No. 3, pp287-289, 1996.

[8] Moore, Brian C. M. (ed.) *Hearing.* Toronto: Academic Press, 1995.

[9] Olson, Harry F. *Music, Physics and Engineering.* New York: Dover Publications, 1967.

[10] Piszczalski, Martin. *A Computational Model of Music Transcription.* PhD Thesis, University of Michigan, 1986.

[11] Quirós, Francisco J. and Enríquez, Pablo F-C. Real-Time, Loose-Harmonic Matching Fundamental Frequency Estimation for Musical Signals. IEEE-ICASSP 1994, Vol. II, pp221-224.

[12] Steedman, Mark. The well-tempered computer. Phil. Trans. R. Soc. Lond. A., Vol. 349, pp115-131, 1994.