

University of Arkansas, Fayetteville

ScholarWorks@UARK

Computer Science and Computer Engineering
Undergraduate Honors Theses

Computer Science and Computer Engineering

5-2008

Pitch correction on the human voice

Suzanne Ownbey

University of Arkansas, Fayetteville

Follow this and additional works at: <https://scholarworks.uark.edu/csceuht>



Part of the [Computer Engineering Commons](#), and the [Graphics and Human Computer Interfaces Commons](#)

Citation

Ownbey, S. (2008). Pitch correction on the human voice. *Computer Science and Computer Engineering Undergraduate Honors Theses* Retrieved from <https://scholarworks.uark.edu/csceuht/19>

This Thesis is brought to you for free and open access by the Computer Science and Computer Engineering at ScholarWorks@UARK. It has been accepted for inclusion in Computer Science and Computer Engineering Undergraduate Honors Theses by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu.

Pitch Correction on the Human Voice

An Honors Thesis submitted in partial fulfillment of the requirements
for Honors Studies in Computer Science

By
Suzanne Ownbey

Spring 2008
Computer Science
College of Engineering
The University of Arkansas

Introduction

Singing is an excellent way to achieve fame, granted one can sing on key. Alternatively, a person who does not inherently have this ability or want to spend the time practicing to achieve it could use a programmatic solution to pitch correction. Proprietary products already exist that can correct the pitch of a singer, but the average mediocre singer may not be willing or able to purchase them.

The focus of this research is to use the basic frequency altering algorithm of resampling, which alters the length of the wave to correct the pitch of a note. Small parts of the wave will be resampled to a harmonic pitch to keep the wave at its original length. (See Figure 1.) If a wave is resampled to a lower pitch, the length will be increased, and the harmonic pitch will be higher so it can shrink parts of the wave. If a wave is raised in pitch, the length will decrease and the harmonic pitch will be lower to stretch the length make to its original size.

Figure 1 a. Shows a wave whose pitch needs to be raised to be in tune.

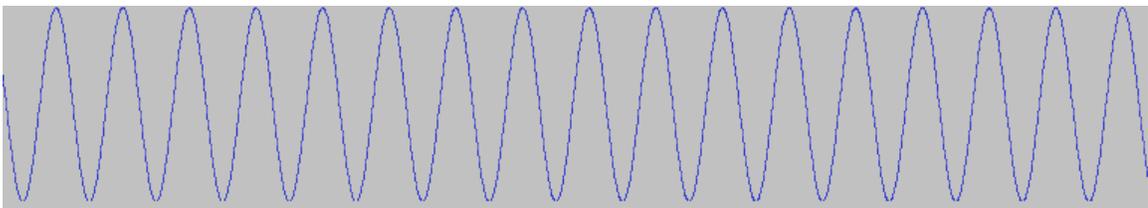
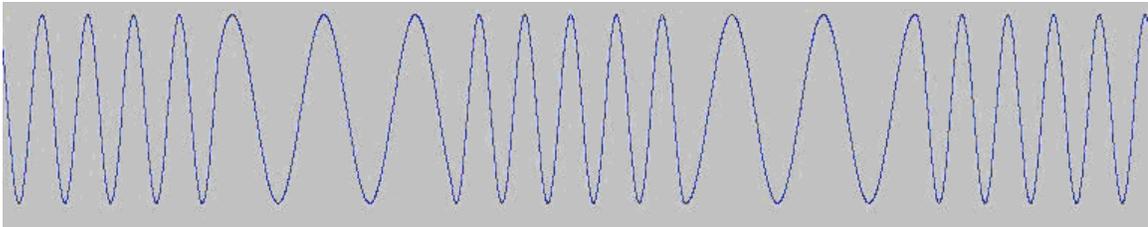


Figure 1 b. shows what Figure 1 a. would look like after it had been resampled to a higher frequency and parts resampled to a lower harmonic frequency.



Background

Little academic research has been published on altering frequency of sounds waves. One algorithm, named Pitch-Synchronous Overlap and Add or PSOLA^[2], corrects the length changing problem of resampling. It sections the wave and adds or removes parts to keep the length the same as the original, but this can cause discontinuities and is complicated.

At the core of this research is the resampling algorithm. Resampling is a process by which a wave's frequency and thus the pitch can be altered. Digital wave files consist of points of data, samples, across the wave taken at a standard time interval know as the sample rate. By removing some of the samples in the file and keeping the time step constant, the frequency of the wave can be increased. To lower the frequency, the samples can be repeated or synthetically reproduced by interpolation to increase the number of samples. A resampling factor is used to determine which samples to keep, skip, or interpolate and is found by dividing the desired frequency of the note by the original frequency.

Equation 1. The resampling factor, F

$$F = \frac{\textit{desired frequency}}{\textit{current frequency}}$$

This factor is then used as a guide to iterate through the samples. Below is the basic resampling algorithm.

Algorithm 1. Basic Resampling

```

loop through samples until end
  newSample.add(samples[index])
  index = index + factor
end loop

```

For example, a resampling factor of two will remove half of the samples by removing every other one, which causes the same number of peaks to be present but in half the amount of time. So a factor of two will double the frequency causing the pitch to be raised an octave. To raise a pitch less than an octave, the factor has to be greater than one and less than two. If the resampling factor is not an integer, then samples cannot merely be left out of the new wave. Instead, points have to be interpolated for the fractional parts in order to properly change the frequency.

This process of resampling will either decrease or increase the duration of the wave in proportion to the resampling factor, which leads to discontinuity when trying to alter the pitch of a note within a song. The length of the wave after resampling is the original length divided by the resampling factor (Equation 2.) If the resampling factor is two, for example, then the length will be cut in half.

Equation 2.

$$LengthAfter = \frac{LengthBefore}{Factor}$$

To find the resampling factor, it is necessary to know the pitch of the given wave. Several algorithms exist that accomplish this, but none are perfect. The flaws in these algorithms will be ignored and encapsulated, so that if in the future, better algorithms are found, they can replace the algorithms used in this research. The YIN algorithm is one such algorithm.^[1] The accuracy of YIN decreases with the clarity of the sound, so initial tests will use perfect sinusoidal waves to remove error from determining the frequency.

To choose a pitch to resample the intermediary sections to, the frequencies of the corrected pitch and the length correcting pitch must be harmonious. In order to find a harmonious pitch, a little music theory is needed. According to the *Oxford Companion to*

Music, an interval in music theory is “the distance in pitch between two notes.”^[3] This distance can be described using the diatonic scale or frequencies. The three intervals that are said to sound the best to the ear are known as perfect fourth, perfect fifth, and an octave. To find a note that is an octave higher than a base note, the frequency has to be doubled so the ratio between pitches that are an octave apart is 1:2.^[4] The pitch ratio of a perfect fifth is 2:3 and a perfect fourth is 3:4.^[4]

Experiment

Overview

In order to raise the pitch of a note, the data will be resampled to a corrected frequency for most of the wave and to a harmonic frequency periodically to lengthen or shorten the wave. The factor used to create a harmonious pitch is referred to as the secondary factor and is harmonious so that the change from between pitches will be less detectable by the listener. How long and how often sections are resampled using the second factor will be determined by how much the wave needs to be lengthened. The length and rate of recurrence of these sections are bounded by the necessity of keeping these sections undetectable, which in turn puts a lower limit to the length of the wave that can be corrected using this algorithm, because a short wave would not be able to hide the harmonious sections of the wave. A bound also exists on how much the algorithm can change the frequency of a note. Since this algorithm is only trying to help mediocre singers rather than tone deaf ones, the bound is not an issue, because mediocre singers only need a small amount of help with their pitches.

Another issue this algorithm will correct is the transition from one note to another in the same sound file. If the waveforms of two consecutive notes do not match up, an inconsistency in sound can occur that the listener would identify as synthetic. Since this algorithm will treat each note separately, no transition correction can occur. Therefore, I will not change the existing transitions from one note to the next by not resampling a small portion of the beginning and end of each note.

Figure 2 is an example of resampling and pitch correction preformed by this algorithm. The wave in *Figure 2 a.* has a frequency of 400 cycles per second for five one hundredths of a second, which barely sounds any longer than a click. This wave is run through the resampling algorithm to raise its pitch to 440 cycles per second and the result is shown in *Figure 2 b.* Note that because the original wave was raise in pitch, the resulting wave is shorter. The longer the original wave was the greater this difference would have been. The wave in *Figure 2 c.* is the result from running it through the length correction algorithm. Notice that it is not only the same length as the original, but also the amplitude of the end of the waves is the same. This is important if this wave is really part of a larger wave so that the transition from the corrected section to the rest is not audible. Also, two sections have been noticeably been resampled using a factor of three fourths causing the frequency of this sections to be 330 cycles per second.

Figure 2 a. Original Wave, 400 cycles per second, 0.05 seconds

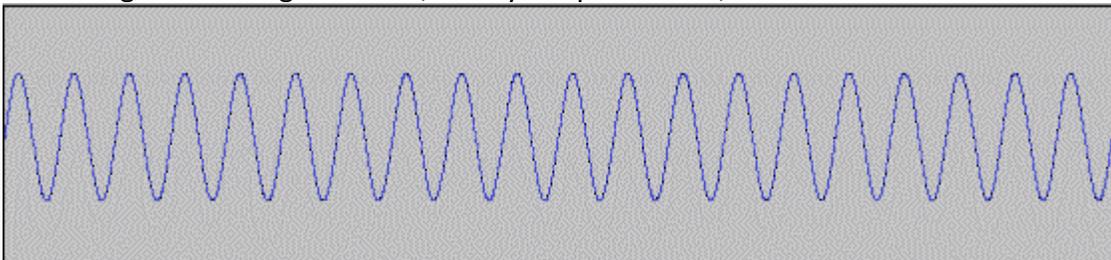


Figure 2 b. The wave in a. resampled to 440 cycles per second

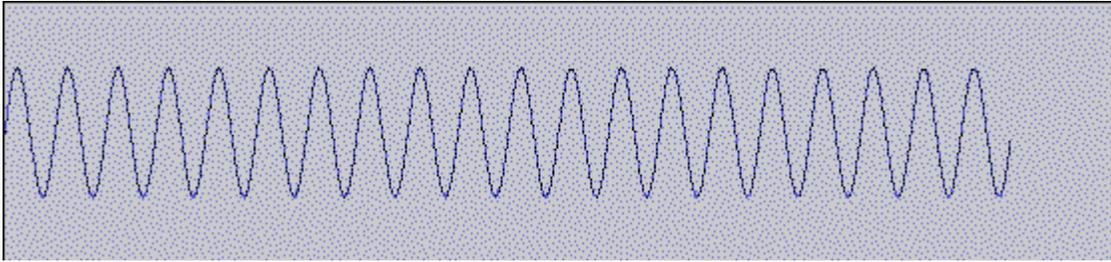
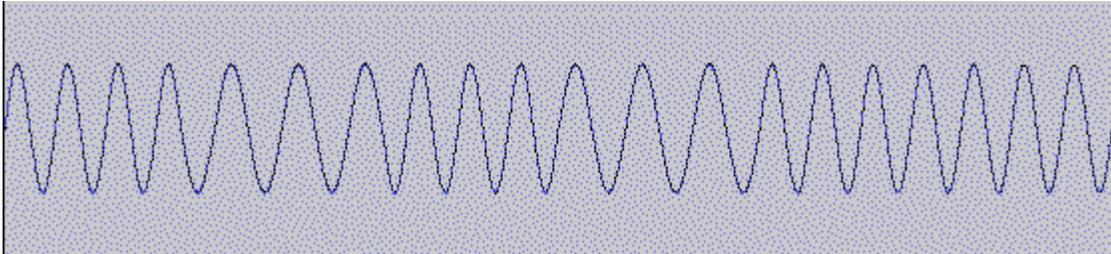


Figure 3 b. The wave in b, with sections resampled to 330 cycles per second



The implementation of this algorithm can be divided into two sequential parts. First the number of samples to resample at the secondary factor must be found. Then, the problem of how to disperse the samples to be resampled using the secondary factor must be tackled. A decision was also made to correct the pitch of the note first and then resample parts of the wave using the secondary factor, which will be computational more expensive than if all of the resampling was done a one time, but will still show whether this technique will work or not. Within the second step, the secondary factor must be calculated to change the pitch to a harmonious frequency. To resample parts of a wave that has all ready been resampled to the correct pitch, the resampling factor will be one of the pitch ratios of the three intervals said to sound best to the ear: perfect fourth, perfect fifth and the octave.^[4] If length needs to be added to the wave then the secondary factor is just the fraction of these ratios; however, when length needs to be subtracted, the inverse of the fractions is used.

Number of Samples to Resample

The first challenge is to determine how many samples to resample using the secondary factor to restore the wave to its original length. The length of wave after resampling, L_a , is equal to the length of the wave before resampling, L_b , divided by the resampling factor, F as shown in Equation 3.

Equation 3. The length after resampling

$$L_a = \frac{L_b}{F}$$

This equation was used as the basis in creating an equation for the number of samples to be resampled at the secondary factor, which is referred to as R in this research. In the above equation, R would be the L_b , because it is the length of the samples before resampling. The factor, F , is the secondary resampling factor, F_2 , which is one of the harmonious fractions. The length after resampling needs to make up the difference between the original length, L_o , and the corrected length, L_c ; therefore, L_a needs to be the difference of L_o and L_c . Putting this all together, Equation 4 is created.

Equation 4. First attempt to find R , the number of samples to resample at the secondary factor, F_2

$$L_o - L_c = \frac{R}{F_2}$$

But this does not exactly correct the length, because part of the original wave has been resampled by F_2 and must be accounted for in the length after resampling. So the length after resample must include the number resample by F_2 .

Equation 5. The addition of R to the left side of Equation 4

$$L_o - L_c + R = \frac{R}{F_2}$$

Equation 6. The result after solving Equation 5 for R

$$R = \frac{F_2(L_o - L_c)}{(1 - F_2)}$$

As shown by Equation 6, it is likely that R will not be an integer, which is a problem because it is hard, if not impossible, to resample a fraction of a sample. Also the rounding error caused by division needs to be minimized. In order to accomplish this, R needs to be guaranteed an integer. Using the knowledge that F_2 is some integer, n, divided by another integer, d, R can be insured to be an integer. These integers n and d must be further restricted to be consecutive so that the absolute value of their difference is always one. All three ratios used for the secondary factor adhere to this construct. Applying both of these restrictions, *Equation 6*. can be simplified as shown by *Equation 7*.

Equation 7. The simplification of Equation 6.

$$R = \frac{\frac{n}{d}(L_o - L_c)}{1 - \frac{n}{d}} = \frac{\frac{n}{d}(L_o - L_c)}{\pm \frac{1}{d}} = \frac{n}{d}(L_o - L_c)(\pm d)$$

$$R = \pm n(L_o - L_c)$$

Whether n is positive or negative is depended on whether one minus the resampling factor is positive or negative. If the resampling factor is greater than one, then the result will be a negative number. In this case, the wave must need to be shortened and therefore, the difference of the original length and the correct length will be negative, which would make R a positive number. If the resampling factor is less than one, then n will be a positive number and the difference of the lengths will be positive as well. (The equation is discontinuous when the resampling factor is one. A resample factor of one would never be use because there is never a need to resample a wave to its

current frequency.) The plus or minus can be dropped and instead the absolute value be taken, because R will be positive in either case.

Equation 8. The number of samples to resampling using the secondary factor

$$R = |n(L_o - L_c)|$$

Distribution of Secondary Resample

Now that the number of samples to resample can be found using Equation 8, it is time to determine how to distribute them throughout the wave. Two different approaches were used; one was an equal distribution of sections of R and a distribution R by a harmonious frequency. In general, the approach of the equal distribution was to divide R into sections that were some multiple of the wavelength of the desired frequency and spread these sections evenly throughout the wave. Since R most likely will not be divisible by the length of the sections, the last section would contain the remainder of this division. This could cause the last section create a jerk of in the overall sound. How severe the jerk is related to the difference of lengths of this section and the rest of the sections. This will be kept in mind when the experimental waves are reviewed and could be corrected with further research.

So to find the lengths of these sections, the wavelength of the wave must be found by dividing the sample rate, by the desired frequency of the note. The wavelength was then multiplied by some integer and the result was used as the length of the sections. Whole wavelengths were used to keep the integrity of the sound. Several different integers were tested to see the effect they would have on the overall sound. To find the number of sections needed, R was divided by the length of the sections. The remainder

of this division is put into an extra section at the end section, so the number of sections needs to be increased by one.

Algorithm 2. Equal Sectioning of secondary sections

$$\begin{aligned} \square &= \text{sample_rate} / \text{desired_frequency} \\ \text{Section_Length} &= \square * N \\ \text{Num_Sections} &= (R / \text{Section_Length}) + 1 \\ \text{Last_Section_Length} &= R \% \text{Section_Length} \end{aligned}$$

The second algorithm created distributed the sections resampled by some harmonious frequency, which will be referred to as the meta-frequency. The meta-frequency measures how often a section of the harmonious pitch occurs. It must be significantly less than that of the either the harmonious pitch or the corrected pitch to preserve any of the integrity of the sound. In other words, the number of times a section occurs in a second is must be smaller then the number of times a wavelength of one of the frequencies occurs in a second. The idea is to create a secondary section and a normal section that form one cycle that is the wavelength of the meta-frequency.

Equation 9. The equation for the length in samples of a meta-cycle where sectionL is equal to the length of the secondary section and skipL is equal to the length of the normal section.

$$\text{Meta - cycle} = \text{sectionL} + \text{skipL}$$

The meta-frequency needs to be found using a combination of the three intervals that are pleasant to the ear, most likely either perfect fourth or fifth several octaves lower.

Next, the lengths of the secondary section and the normal section need to be found. To find them, an equation using the meta-frequency and these two variables was set up, Equation 10. The lengths of these to sections are also dependent on the number of samples to resample at the secondary rate, so Equation 11 was set up to allow the variables to be solved. This equation was solved for the skip length and substituted for

into Equation 10, which lead to Equation 12. (Equations 10-13 are only for the case of raisings the pitch.)

Equation 10. Meta-frequency using the section length and the skip length

$$meta - F = \frac{sampleRate}{(sectionL + skipL)}$$

Equation 11. Number of meta-cycles in the wave where L is the length of the wave after primary resampling

$$numberCycles = \frac{R}{sectionL} = \frac{R + L}{sectionL + skipL} \Rightarrow$$

$$skipL = \frac{L * sectionL}{R}$$

Equation 12. Section Length in terms of the constants R, sample rate, meta-frequency and length after primary resampling.

$$meta - F = \frac{sampleRate}{sectionL + L * \frac{sectionL}{R}} = \frac{sampleRate}{sectionL * (1 + \frac{L}{R})}$$

$$= \frac{sampleRate}{sectionL * \frac{(R + L)}{R}} = R * \frac{sampleRate}{sectionL * (R + L)} \Rightarrow$$

$$sectionL = \frac{R * sampleRate}{meta - F(R + L)}$$

Equation 13. Skip length in terms of the constants R, Sample rate, meta-frequency and length after primary resampling

$$skipL = \frac{L}{R} * \frac{R * sampleRate}{meta - F(R + L)} = \frac{L * sampleRate}{meta - F(R + L)}$$

Testing and Results

In order to ignore the problem of finding the pitch of a test wave, Audacity, an open source sound editing software, was used to create perfect sinusoidal waves of a know frequency. The pitch of middle A, which has a frequency of 440 cycles per second,

was used as the target pitch for testing. The incorrect frequencies used to test the algorithm ranged from 430 to 450 cycles per second. A difference of ten cycles per second would be equivalent to a little better than average, but still mediocre singer. Also, each one was tested with a perfect fifth, two thirds, and perfect fourth, three fourths as the harmonies pitch. The octave was used a little at first, but was too great of a change to be practical for this research.

The algorithm hinges on the ability to properly find how many samples to resample at the secondary factor, so the first tests tested the accuracy of calculating the number of samples to resample using the secondary factor. To do this, R was found, then that number of samples was resampled in a one lump at the end of the wave and this length was compared to that of the input wave. For all test cases, when R samples were resampled, the wave was returned to its original length. Equation 8 did work. The estimation involved when resampling using a fractional part did not affect the outcome. The estimation comes into play because the resampling loop is to stop when a double is past an integer. This makes it hard to be sure that the correct number of samples was resampled and could be more of an issue with different frequencies.

Next, the sectioning part of the algorithm was implemented. This part was a little harder to implement because a formula could not just be derived. Instead, the manipulation of data structures, namely arrays had to be handled. A universal index was used to keep track of the current position in the original wave and put both the resampled and copied portions into a new array. By copying into a new array, the change in length from resampling did not affect the indexes in the original wave. Basically, the algorithm copies the wave at the index and increments the index for a set skip period and then

resamples for the length of the secondary resample section and then increments the index the right amount of times.

Algorithm 3. Equal Distribution of the secondary sections

Note that Algorithm 2. finds some of the values in this algorithm.

```

skip = (wavsize-R)/(numSections+1)
//we need 1 more skip group than number sections
Loop for each numSection
    Copy samples from index to index + skip
Resample from index to sectionLength +index
    Index +=sectionLength
End Loop

```

Algorithm 3 assumes there is no remainder when R is divided by the section length to determine the number of sections. The algorithm must resample exactly the number of samples determined by R, so Algorithm 3 must be added to include any leftover samples.

Algorithm 3. Sectioning the wave for the secondary resample.

```

Copy samples from index to index + skip //
increments index
Resample from index to (R % sectionLength) +index
Index +=R % sectionLength
Copy samples from index to end of wav

```

When the secondary resample portion was equally distributed, the length of the original was not perfectly maintained. The algorithm made the corrected wave longer by less than one thousandth of a percent. The longest this excess was observed was %0.000227 which would translate into approximately 4.53×10^{-4} seconds with the sample rate at 44100 cycles per second. This duration makes the difference irrelevant because it is not audible. This did not show up when all of the secondary resampling was done in one section, because the rounding error from resampling only occurs once at the end of the resampling. When the sections were dispersed throughout the wave, the rounding error was present every time a section was resampled. This error was smaller

when the perfect fourth, three fourths, was used compared to the perfect fifth, three thirds. The perfect fifth showed less of this error because it is a rational number, which would make any rounding error smaller, and the perfect fourth is an irrational number.

Success was found for the restoring the wave to its original length; however, the sound integrity was not. An attempt was made to calibrate this difference by changing the length of the secondary sections. The larger the sections are the more and more audible they are, but they also become more natural to the ear. The smaller the sections are the less audible they are, but the overall sound is more synthetic. The smaller sections sound more like a click, and the longer sections sound more like a siren. Also, the less the wave needs to be corrected in the first place, the less often a sound disruption occurs, but none of the waves test sounded natural.

To test the harmonious distribution of the secondary sections, first the concept of a meta-frequency was tested by arranging a wave with a meta-cycle consisting of one wavelength of with a frequency of the perfect fourth followed by three wave lengths of the base frequency. This uses the ratio of perfect fourth, three sections of base frequency for every four wavelengths. The result is almost a pleasant sound that is like that of a string instrument. The overall frequency, however, is several octaves lower than that of the desired frequency for the note. This overall frequency is known as the fundamental frequency of a note and occurs when the note is not purely a single frequency but rather some combination of sound waves.^[4] The frequency of a human voice is actually the fundamental frequency of the sound, because the sound is created by more than just a single piece of anatomy vibrating. Although this researched finally produce a wave that

sounded descent, the fact that the sound was not at the correct frequency was discouraging.

In order to distribute the sections in a harmonious manner, the equations for creating a meta-frequency, *Equations 12 and 13*, were used in *Algorithm 2* to replace the section length and the skip length. Meta-frequency was set to fifty-five, 110, 220, 330, 440 cycles per second because each is a different octave or an interval of perfect fourth for the desired frequency. In each case, the skip length was found to have a ratio of three fourths to the total number of samples in the meta-cycle. Also in each case, the sound produce was a perfect fourth lower than the frequency of 440 cycles per second. The only difference that occurred with the use of different meta-frequency was how pure the sound was. A meta frequency of 440 cycles per second sounded the least synthetic.

Because of this result, the secondary section needs to be the length of some multiple of the wavelength. Using the ratios of the best sounding intervals, the length of the skip section should also be a multiple of same wavelength. In order to test this, the wavelength was found using the equation in *Algorithm 2* then various ratios derived from the best sound ones were used. *Table 1.* shows the ratios that were tested.

Table 1. Combinations of section lengths and skip lengths tested and the ratios they create.

sectionL	skipL	Ratio
1	3	3:4
1	7	7:8
2	6	6:8
2	14	14:16
3	5	5:8
3	9	9:12
3	11	3:11
4	13	4:13
4	28	4:28

5	15	5:15
---	----	------

None of these combinations created a sound that was the target frequency and did not sound synthetic. The sounds created either sounded like the equal distribution of the secondary chunks or sounded like the wrong frequency.

Conclusion

The testing showed that none of the approaches in this research work with perfect sinusoidal waves; however, the mediocre singer does not sing perfect sinusoidal waves. Instead, they produce some combination of vibrations that are rougher than the sinusoidal waves. This roughness could hide the synthetic sounds produced by these algorithms. Research using a human's voice needs to be done before a conclusive assessment of the algorithms laid out can be preformed. The poor mediocre singer could still be helped.

References

- [1] A. de Cheveigne. "YIN, a fundamental frequency estimator for speech and music," *Journal of the Acoustical Society of America* 111 (4), April 2002, pp.1917-1930.
- [2] H. Valbret , E. Moulines , J. P. Tubach, "Voice transformation using PSOLA technique", *Speech Communication*, v.11 n.2-3, p.175-187, June 1992
- [3] P. Scholes, et al., "Interval." *The Oxford Companion to Music*. Ed. Alison Latham. *Oxford Music Online*, 16 Apr. 2008. <<http://0-www.oxfordmusiconline.com.library.uark.edu:80/subscriber/article/opr/t114/e3448>>.
- [4] John Borwick,. "Acoustics." *The Oxford Companion to Music*, Ed. Alison Latham. *Oxford Music Online*. 16 Apr. 2008. <<http://0-www.oxfordmusiconline.com.library.uark.edu/subscriber/article/opr/t114/e53>>.