ISSN: 2586-761x (Print)
ISSN: 2635-5817 (Online)
Vol. 01, No. 02, December 2018
International Journal of Advanced Social Sciences
Source: http://ictaes.org
Manuscript received: November 20, 2018 ; Revised: December 20, 2018 ; Accepted: December 25, 2018

Effects of Auto Tuning and Pitch Normalization on Query by Humming

Prasanna Koirala[1], Manoj Chapagain[2], Nishan Pantha[3], Nanda Bikram Adhikar [4]

Department of Electronics and Computer Engineering, Institute Of Engineering, Pulchowk Campus Pulchowk, Lalitpur, Nepal

pkoiralap@gmail.com[1], chapagainmanoj35@gmail.com[2], nishanpantha@gmail.com[3], adhikari@ioe.edu.np[4]

Abstract

Query by Humming (QbH) is a method used for the identification of songs on the basis of hum. There are many factors such as pitch, tempo, sound coefficients, audio instruments, etc. that play part in the identification process. However, pitch is the most significant factor as identified in previous works in this area. Pitch(y) vs time(x) signal of the hummed tune is calculated, which is then compared to pitch(y) vs time(x) signals of the original songs stored in the database. The signals are compared using Fast Dynamic Time Warping (FDTW). This research compares the accuracy of a fundamental QbH system when the signals are auto tuned and normalized over when they are not. Auto tuning process scales the pitch to the nearest absolute pitch value. Normalization process uses mean pitch and standard deviation throughout the signal to scale multiple similar signals to the nearest range. The obtained results over auto tuned pitches have a significant increase in the overall accuracy. Normalization process however showed increase in accuracy only when the number of songs in the database were high.

Keywords*: Query by Humming, Pitch, Fast Dynamic Time Warping, Normalization, Auto Tuning*

1. INTRODUCTION

As the number of multimedia information increased over time; mainly due to the internet and easy accessibility and sharing platforms, the search for a proper way of retrieving that information [1] kept increasing as well. Query by Humming (QbH) is one of the finds of such searches. Often times human brain remembers the melody or tune of the song, an iconic guitar piece or even a catchy chorus but not the title or artist's name by which the one could search for the actual song. QbH systems enable users to search for the song on the basis of tune hummed by the user. These systems are very convenient to use and are constantly evolving to be more and more accurate over time.

A conventional QbH system extracts the pitch of the hummed tune and compares it over the pitches of the saved original tunes that already exist in the database. This research uses Aubio [2]; an open source tool that lets users extract the pitch of a audio signal. These extracted pitches are then compared to the pitches present in the database using FDTW [3].

Corresponding author : Prasanna Koirala
Author's affiliation : *Department of Electronics and Computer Engineering*
Email: *pkoiralap@gmail.com*

Pitch being the most important part of the entire comparison process needs to be further processed for greater accuracy of the system.

2. METHODOLOGY

The database is populated using tunes that replicate the song's identical parts, played on a standard piano. The pitch data is incrementally added to the database manually. A total of 35 Nepali songs were populated in the database. The audio input was taken from smartphone devices.

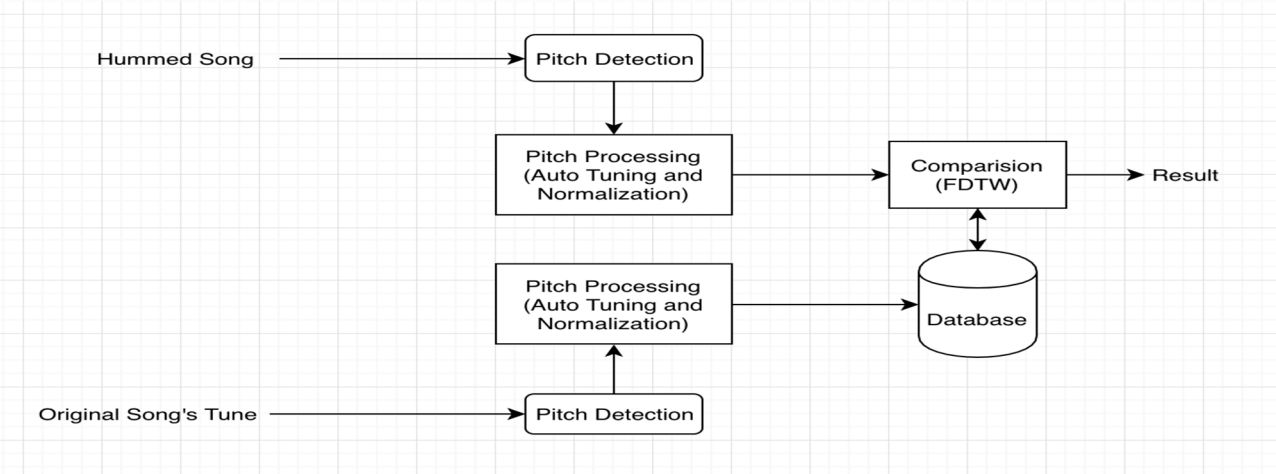

Figure 1. Query by Humming

2.1 PITCH DETECTION

Pitch Detection is done via AubioPitch [2]

aubiopitch [**[-i]** *source*] [**-o** *sink*] [**-r** *rate*] [**-B** *win*] [**-H** *hop*] [**-p** *method*] [**-u** *unit*] [**-l** *thres*] [**-s** *sil*] [**-f**] [**-v**] [**-h**] [**-j**]

2.2 AUTO TUNING

A list of standard pitch values is first stored in an array **S**. If the pitch value at any point of time **t** is **P(t)**, auto tuned value **A(t)** is obtained by the equation:

$$index = i : \ |P(t) - S[i]| \ \ is \ \ minimum \ \ \forall \ \ i \in [0, length(S) - 1] \qquad (I)$$

$$A(t) = \ \ S[index] \qquad (II)$$

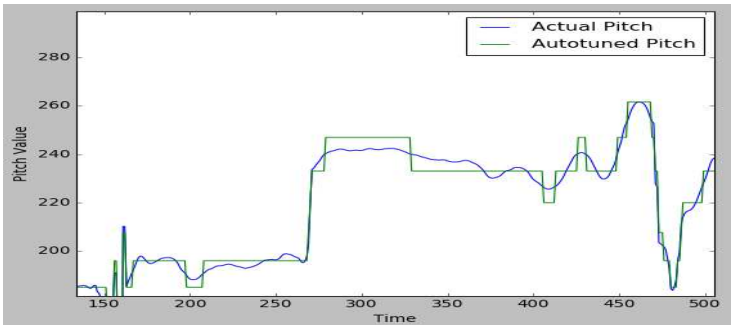

Figure 2. Auto Tuning

2.3 NORMALIZATION

The user may not hum on exactly the required or intended pitch. This may be because of lack of vocal trainings. Normalization process, scales both the original tune and the hummed tune to similar proximities. Comparing normalized tunes then result in lower distance value than non-normalized signals. Normalization is done using mean and standard deviation. The formula for Normalization is:

$$N = \frac{Mean(X) - X}{Std(X)} \qquad (III)$$

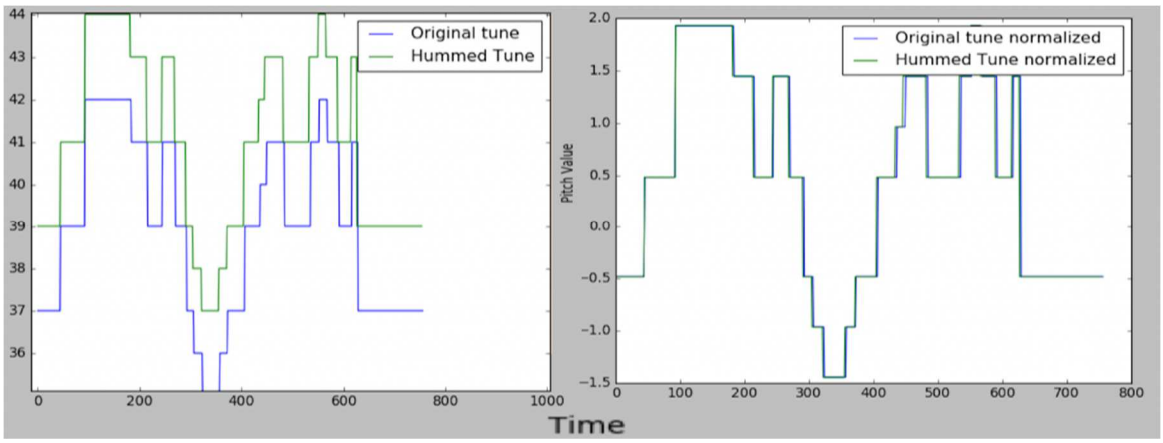

Figure 3. Un-normalization vs Normalization of tunes

2.3 COMPARISON

Fast Dynamic Time Warping (FDTW) [3] is used for comparing two pitch signals with each others. A comparison yields in an overall distance output. The lesser the distance between two signals, the better the match.

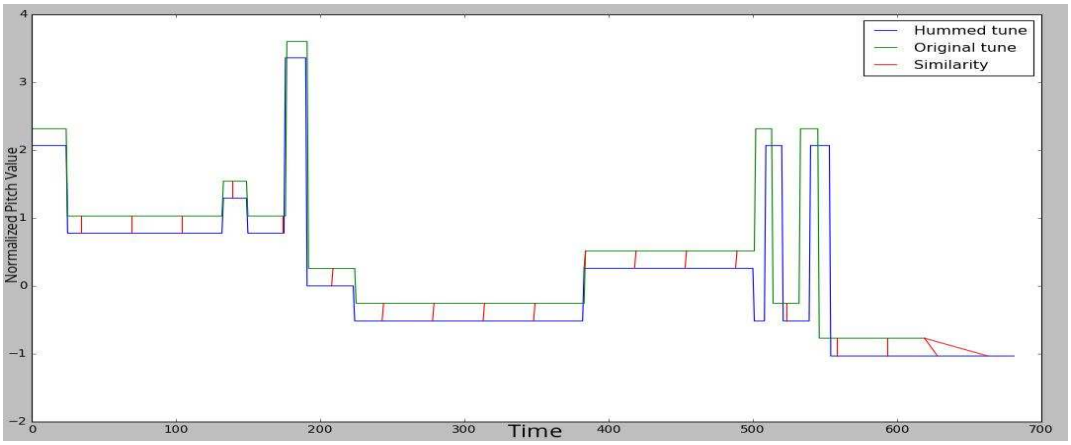

Figure 4. Comparison between a hummed tune and an original tune

3. RESULTS

The experimentation was done with inputs from three users. Each hum was recorded with audio recorder found on smartphones. The results presented are an average of 50 experiments done on each category. The results from individual users were then averaged to obtain the final result.

Table 1. Accuracy without auto tune and normalization

Song number	Song in Top 5	Song in Top 3	Exact Match
10	80.2%	75.31%	68.15%
15	77.72%	69.52%	60.78%
20	71.1%	59.91%	47.24%
25	60.9%	48.24%	35.16%
30	50.8%	32.31%	21.71%
35	43.63%	26.7%	17.41%

The accuracy of the system when the tunes are not auto tuned is drastically low. This is due to the fact that the vocal skills of the users vary between a trained and non-trained vocalist. A more vocally trained user could increase the accuracy substantially.

Table 2. Accuracy when tunes are not normalized(auto-tuned)

Song number	Song in Top 5	Song in Top 3	Exact Match
10	93.12%	91.13%	89.91%
15	90.32%	89.14%	83.25%
20	85.19%	83.93%	78.96%
25	84.2%	77.21%	71.58%
30	84.12%	73.73%	66.9%
35	83.84%	70.1%	60.49%

Auto tuning however increases the hit rate by at most 600%. Most the errors of humming are mitigated by auto tuning process. Hence even a user who is not trained vocally would be providing pitch data that is rather accurate.

Table 3. Accuracy when tunes are normalized(auto-tuned)

Song number	Song in Top 5	Song in Top 3	Exact Match
10	99.83%	97.73%	90.81%
15	95.18%	95.26%	84.5%
20	92.34%	89.79%	82.24%
25	91.14%	88.12%	79.23%
30	90.05%	85.26%	75.35%
35	90.02%	84.87%	73.17%

Normalization process increases the accuracy even more. While it does not seem to have a huge effect on the overall accuracy of the system for lesser number of songs in the database, Normalization has a substantial effect when the number of songs in the database increase

In all three cases a rather higher value of accuracy is obtained for Song in top 5 and Song in top 3. This could be because the precision scale of the match decreases as we increase the range of the output and also because of the type of songs present in the database. It is entirely possible that songs with similar pitch values exist on the

database.

The overall accuracy of the system under all three conditions, seems to be decreasing linearly with the increase in number of songs in the database.

4. DISCUSSIONS

Previous works have used segmentation and signal stretching/contraction methods to improve the accuracy of the QbH systems. Yang et al.[4] present the method of segmentation of nearly clustered pitch values along with anomaly removal and further use signal stretching and contraction with multiple scale factors. The approach increases the accuracy of the system. However, the method is computationally expensive due to multiple comparisons for a single signal. Ghias et al.[5] use relative pitch information for identification of the song. Relative pitches are represented as S, U or D meaning same, up and down respectively. This method relies on the fact that any pitch higher than the current pitch is considered a U and lower considered a D. However, for accurate comparison of songs, a more sophisticated model of pitch representation is required.

Segmentation averages the signal over a cluster to obtain mean value of that cluster. The mean value is not necessarily the intended pitch value of the hum. Auto tuning however provides the exact intended pitch value and hence improves the accuracy of the system. The hums are not necessarily on the same key as that of the song. Normalization scales down both the song's and the hum's pitch data into similar proximity. Hence, normalization is a novel method that further increases the accuracy of any QbH system.

5. CONCLUSION

In this research, we compared accuracy of a conventional Query by Humming system, when the Pitch is auto tuned and normalized in contrast to when it is not. The comparison was done using FDTW [3] algorithm. The results showed that auto tuning increases the accuracy of the system drastically. Normalization of Pitch however does not have substantial effect on the overall accuracy of the system for lower number of songs in the database and only affects the accuracy of the system when the number of songs in the database increase.

REFERENCE

[1] Müller, M. (2007). Information Retrieval for Music and Motion.

[2] Aubio. https://aubio.org.

[3] Salvador, S., & Chan, P. (2004). FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space.

[4] Yang, J. , Liu, J,. & Zhang, W. (2010). A Fast Query by Humming System Based on Notes. INTERSPEECH.

[5] Ghias, A., Logan, J., Chamberlin, D., & Smith, B.C. (1995). Query by Humming: Musical Information Retrieval in an Audio Database. ACM Multimedia.