INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

SPECTRAL EFFICIENCY WITH TIME & PITCH SCALE MODIFICATION OF SPEECH SIGNAL

Venkatesh Bandla[1*], Prof.Nagajayanthi B[2]
[1*] M.Tech(C.E),SENSE,VIT University Chennai, India
[2]Assistant Professor (SR), SENSE, VIT University Chennai, India
Correspondence Author: bandla.venkatesh2013@vit.ac.in

Keywords: Time-scale modification (TSM), Pitch scale modification (PSM), real-time systems, signals estimation, spectrogram inversion.

Abstract

This paper actualizes a calculation for estimation of signals from brief time extent spectra is presented offering a huge change in quality and effectiveness over current strategies. The essential issue is to address the overlapping frames in the wake of scaling. Another critical issue is to make the calculation computationally helpful for applications. The calculation is parameterized to permit tradeoffs between computational requests and the nature of the sign recreation. The calculation is implemented on sound signals time-scale and pitch scale.

Introduction

The magnitude range of a discrete-time signal X(n) is ordinarily acquired from the Short Time Fourier transform(STFT), which is given as

$$X(mS,\tilde{\omega}) = \sum_{n=-\infty}^{\infty} x(n)\, w(n-mS)e^{-j\tilde{\omega}n} \text{------ } 1$$

Where S is the investigation step size, W is the examination window and M is the record of the edges of the STFT. The time domain signal is especially represented to utilizing STFT and the other way around. The Short Time Fourier Transform Magnitude (STFTM) is given by

$$|X(mS,\tilde{\omega})| = |\sum_{n=-\infty}^{\infty} x(n)w(n-mS)e^{-j\tilde{\omega}\mathrm{n}}| \qquad \text{------ } 2$$

The terms in 2 are same as 1. By combining 1 and 2 the spectral components (Frequency and amplitude) of each frame can be found out.

In numerous signal transforming applications in the wake of preparing the spectral data is lost. The STFTM can't be changed over in time space. Case in point X(n) and –X(n) both are diversified signals of same magnitude. For any arbitrary signal there is no particular approach to change over it back to time area signal. Moreover, in a few applications, spectrogram has just been adjusted (or arbitrary) spectrogram, where the arrangement of covering extent spectra may not be a substantial representation of any genuine esteemed sound signal by any means. In such cases, discover a genuine esteemed sign whose spectrogram is as close as expected to the altered or target spectrogram.

To uproot the covering Mean Square Error(MSE) of the STFTM of actual signal and STFTM of altered sign must diminished. The MSE is given

$$D_M[x(n),x'(n)]=\sum_{m=-\infty}^{\infty}\frac{1}{2\pi} * \int_{\tilde{\omega}=-\pi}^{\pi}[|X(mS,\tilde{\omega})| - |X'(mS,\tilde{\omega})|]^2 d\,\tilde{\omega} \quad \text{---- } 3$$

Griffin and Lim [1] proposed a calculation (from this point forward alluded to as the G&L calculation) iteratively applying a forward and backwards Fourier change to merge toward a period area signal with the fancied range. The Fourier transform is utilized to concentrate the phase from an estimation of the time-space signal. This phase data is then joined with the target magnitude spectrum and is utilized to figure a inverse Fourier transform to produce the following evaluation of the time-area signal.

This paper actualizes the G&L calculation for the real time situation in the beneath segments. Execution pace is an issue with the G&L calculation on the grounds that it requires the reckoning of an extensive number of Fourier transforms for every frame keeping in mind the end goal to accomplish superb reproduction. All the more generally, the classic G&L calculation is unseemly for utilization continuously applications in light of the fact that it requires the magnitude spectra from all frames, past and future, to ensure that the spectral error diminishes monotonically.

A functional calculation is required to satisfy the accompanying necessities.

1. Structural prerequisite: The calculation ought to reproduce casings utilizing just temporal local and potentially past data as opposed to future time data.
2. Computational burden prerequisite: The measure of computation needed to reconstruct the sound signal ought to be sufficiently low to be utilized as a part of real time applications across a wide variety of platforms. .
3. Quality prerequisite: The reproduction ought to be precise if a target signal is known, and free of diversions.
4. Flexibility prerequisite: The reversal calculation ought to reproduce better quality signals with more computational assets.

RTISI algorithm

To satisfy the structural necessity, the signal ought to be built by time sequential order (edge by- edge). In the classic G&L calculation, all the edges are upgraded simultaneously. To satisfy the processing burden prerequisite, the quantity of change cycles must be kept to a minimum.

Assume as of now the first m-1 frames of synthesis signal have been recreated, which is meant as ym-1. Let's consider the problem of generating frame. The signal frames effectively created as of right now are represented in Fig. 1. All of the algorithms use the scaled Hamming window Hamming window.

$$w(n) = \begin{cases} \frac{2\sqrt{S}}{\sqrt{(4a^2+2b^2)L}}\left(a + b\cos\left(2\pi\frac{n}{L}\right)\right), & \text{if } 1 \le n \le L \\ 0, & \text{otherwise} \end{cases} \quad \text{----- } 4$$

where is S the step size between adjacent frames, L is the window length, a = 0.54 and b = -0.46. In our signal reconstruction system, typically L =4S is used, so that the sum of the squares of the overlapping scaled Hamming windows is always 1.

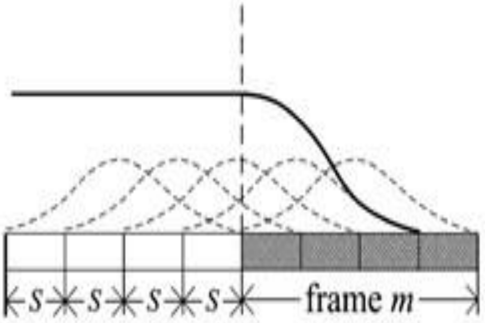

Fig. 1 Illustration of the partially reconstructed frames of signal y(n).

As indicated in Fig. 1, preceding estimation of frame m (for m > 1), the overlap interval is in partially filled by the previous edges. Unless overall indicated, a settled 75% synthesis window covers. So that the mth partially edge originates from the cover included aftereffects of the estimation of the edges m-2,m-3 of y(n), while the final quarter of casing is each of the zero.

To produce an initial evaluation for the periods of frame m, compute the phase of the partially recreated signal utilizing an examination window situated at the partially constructed edge m. This guarantees that even without iterating, the beginning stage phase estimate for frame m will give great stage congruity the partially recreated signal. The Fourier transform of this partially frame is ascertained with the scaled Hamming window. The subsequent stage data is consolidated with the target magnitude range for the M-obliged change step. The Inverse Fourier transform of this new frequency domain signal creates another estimate of frame m.If the target number of iterations has not been reached, frame m is added to the fractional frame ym-1(n)w(n-mS), the window is applied, and afterward the Fourier transform of the windowed summation is ascertained to get another evaluation of the phase. Edge by edge iterative methodology is indicated in figure-(2).

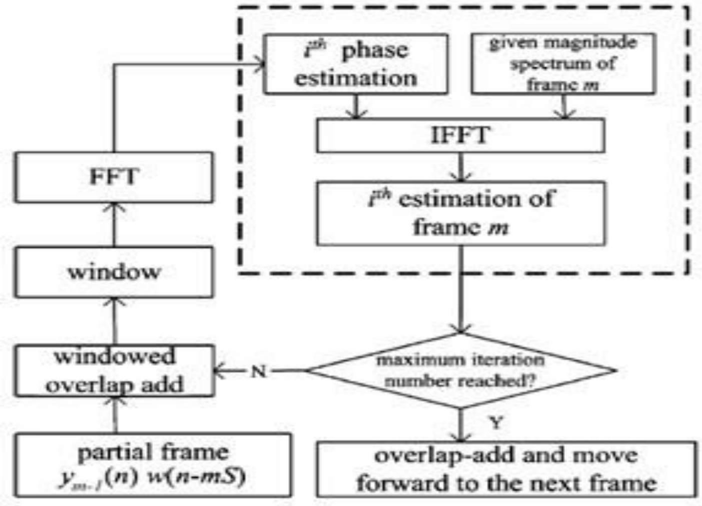
i^{th} phase estimation
given magnitude spectrum of frame m
IFFT
i^{th} estimation of frame m
FFT
window
windowed overlap add
N
maximum iteration number reached?
Y
partial frame $y_{m-1}(n)\, w(n-mS)$
overlap-add and move forward to the next frame

Fig. 2 Frame-by-frame iterative phase estimation Process.

Time scale modification

TSM of signs has long been a subject of enthusiasm for the sound and speech processing areas. A key test in TSM is to change the audio rate, while safeguarding different qualities, for example, pitch and timbre. The time-domain TSM routines function admirably modification factor is near to 1 and when the sign source is monophonic. The RTISI-LA technique functions admirably for both monophonic examples and polyphonic signs.

For an alteration rate a, an examination step size Sa used to get the STFTM and utilize a synthesis step size Ss such that Ss = Sa/ a .The frame lengths in the analysis and synthesis procedure are both L. Here, a settled synthesis step size Ss = L/4 has been utilized, which keeps computational prerequisites steady for different alteration rates. The methodology is indicated in Fig. 3.

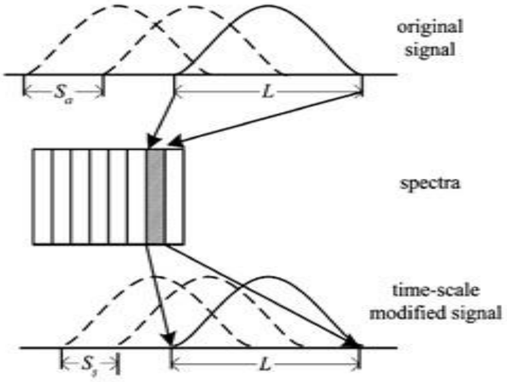

Fig. 3 Time-scale modification in RTISI

The examination step size Sa can be of any arbitrary value. If the examination step size is positive and less than the synthesis step size Ss, the outcome is time-extending, i.e., the reproduced sound plays back more gradually than the first. In the event that the examination step size is bigger than the synthesis step measure, the outcome is time-compression (the recreated sound plays back speedier than the original).The investigation step size can likewise be negative for "reverse" playback, or even zero.

Pitch scale modification

PSM of sound signals is valuable in various applications, for example, multimedia sound signal handling, speech synthesis, vocal identity transformation, and making uncommon sound effects for applications, for example, karaoke. The key is to produce an "extended" or "squashed" range for every frame for upwards or downwards pitch movements, respectively. Modifying the spectrum straightforwardly is hazardous; however, as the the bin spacing for typical edge lengths is very coarse. This makes exact control of pitch shifts about unimaginable.

To reduce this issue and permitting exact pitch shifts, the pitch is shifted by resampling every frame in the time-domain, before processing the STFTM. In the event that the ordinary investigation frame length is L, and the pitch is to be moved upwards by element of q, then for every frame a block of samples L`=qL has been utilized, as demonstrated in Fig. 4(b). Resampling is done in the time-domain to create a casing of frame L as indicated in Fig. 4(c). The target STFTM is then figured on the resulting frame

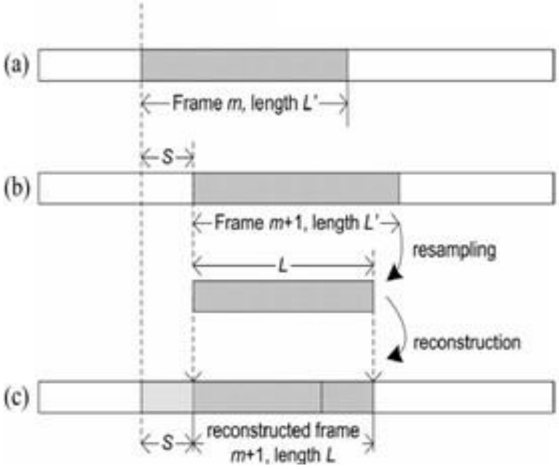

Fig. 4 Pitch -scale modification in RTISI

Results

For evaluation of these TSM and PSM a couple of real time samples has been taken and the time scaling and pitch scaling has been done. The spectrum of the input signal is shown below in fig. 5

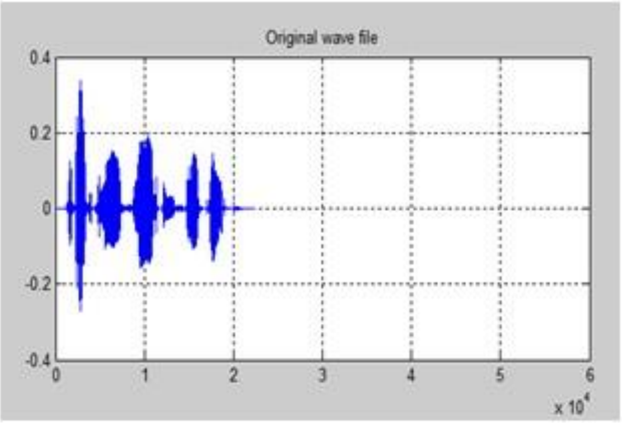

Fig -5 Original Wave file

After the scaling has been performed by a factor of two(*L=2*)

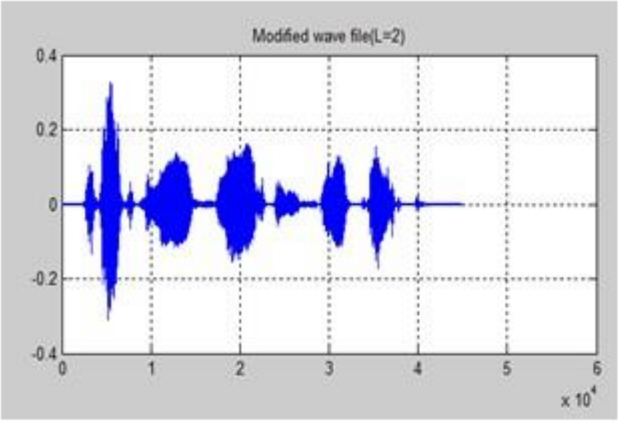

Fig – 6 Time Scaled Wave File (L=2)

As mentioned above the content of the signal should not be varied and if above algorithm is followed it is not changed which can be observed from the power spectral density (PSD) of the input and output wave files.

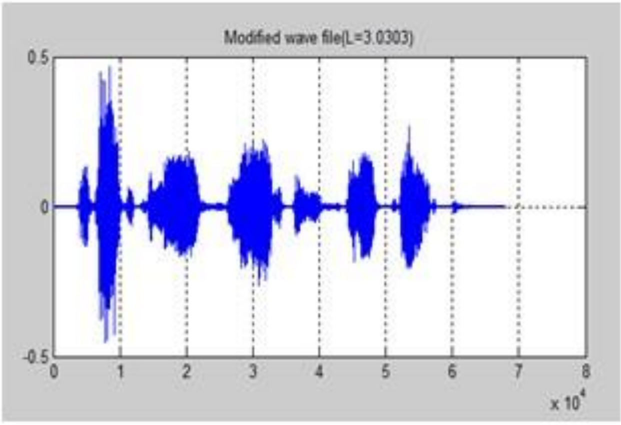

Fig – 7 Time Scaled Wave File (L=3)

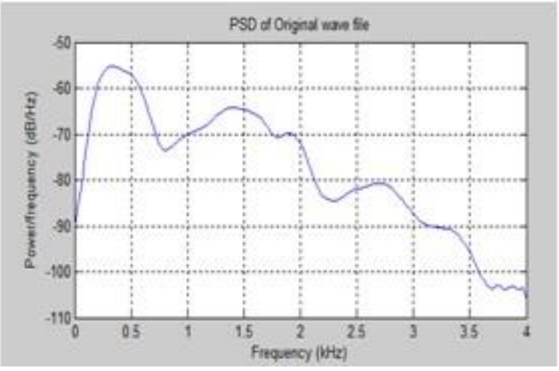

Fig -8 PSD of Original Wave file

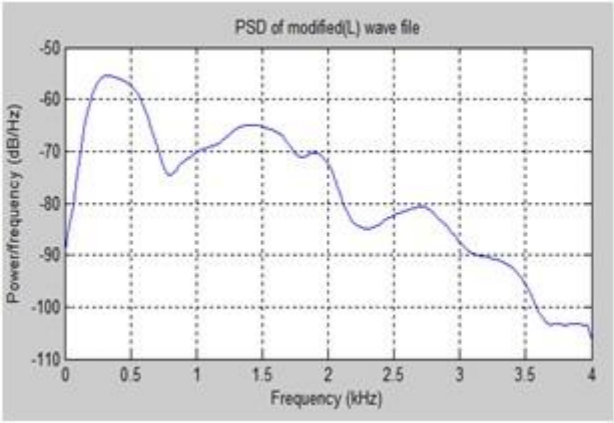

Fig -9 PSD of Modified Wave file

Using the pitch scaling method the pitch scaling had been performed to the same wave file and the output spectrum is shown in figure 10

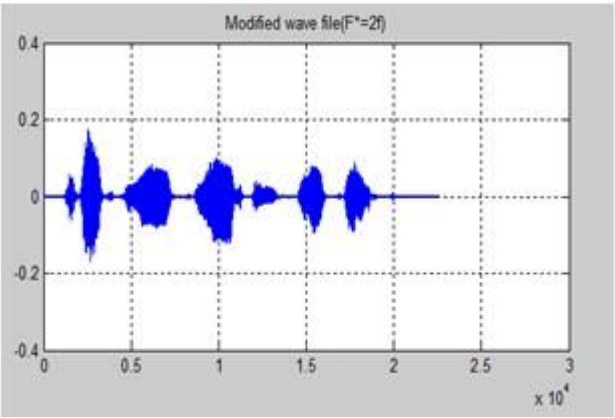

Fig- 10 Output spectrum of PS Modified Wave file

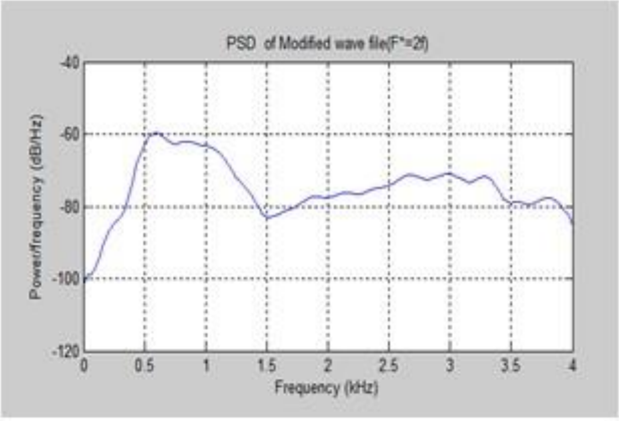

Fig- 11 Power Spectral Density of Pitch Scale Modified Wave file.

Conclusion

By this method the TSM has been performed without any change in the spectral content but a small change in the spectral content has been observed for PSM.

References

1. Improving Time-Scale Modification of Music Signals Using Harmonic-Percussive Separation By Jonathan Driedger, Meinard Müller, and Sebastian Ewert in IEEE SIGNAL PROCESSING LETTERS, VOL. 21, NO. 1, JANUARY 2014
2. Real-Time Signal Estimation From Modified Short-Time Fourier Transform Magnitude Spectra Xinglei Zhu, Gerald T. Beauregard, Member, IEEE, and Lonce L. Wyse. In IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 15, NO. 5, JULY 2007
3. D. W. Griffin and J. S. Lim, “Signal estimation from modified shorttime fourier transform,” IEEE Trans. Acoust., Speech, Signal Process.,vol. ASSP-32, no. 2, pp. 236–243, Apr. 1984.
4. M. H. Hayes, J. S. Lim, and A. V. Oppenheim, “Signal reconstruction from phase or magnitude,” IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-28, no. 6, pp. 610–623, Dec. 1980.
5. G. Michael and M. Porat, “On signal reconstruction from fourier magnitude,” in Proc. 8th IEEE Int. Conf. Electron., Circuits, Syst., Sep. 2001, vol. 3, pp. 1403–1406.