**Paper report**

Submitted in requirement of

DSP-LAB Course

Name- Ankit S Tendulkar

Id-N17400815

Topic: **Phase Vocoder**

Paper: ‘NEW PHASE-VOCODER TECHNIQUES FOR PITCH-SHIFTING, HARMONIZING AND OTHER EXOTIC EFFECTS’ by Jean Laroche, Mark Dolson

**Abstract**

Time scaling and pitch shifting is generally carried out using phase vocoders, which provide excellent quality results. This paper underlines the algorithms for modification of the signal in the frequency spectrum. We can achieve difficult and complex manipulations which are unattainable by simple techniques. There are two stages in the process

1. Peak detection stage
2. Peak shifting stage

**Introduction**

Many of the time scaling applications use phase vocoders. It basically involves breaking the signal into blocks and then changing them into the frequency spectrum using Fourier transform. Amplitude and phase manipulation is carried out in this spectrum and then the blocks are change back to the time domain to get the result. A limitation of this approach is that only linear manipulations can be done by this method. This report discusses two new approaches where complex manipulations can be made in the frequency spectrum. The procedure follows two steps namely.

**Peak detection stage**

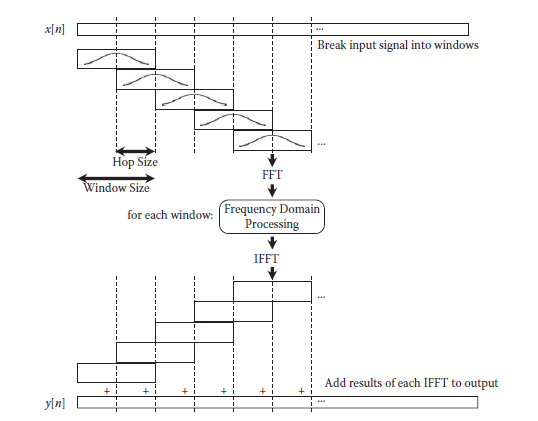
Instead of creating discrete blocks as in case of normal phase vocoders we divide the frequency spectrum of the signal into ‘influence regions’ based on the locations of peaks in the spectrum.

**Peak shifting stage**

These ‘influence regions’ are then shifted to new positions to get the desired effects.

In the usual phase vocoder techniques, we need to use phase unwrapping and instantaneous frequencies to ensure continuity of the phase of the output signal. However, our algorithm requires very simple phase adjustments to keep the phase continuous. Two types of algorithms result, one with integer shifts and other with fractional shifts. The integer shift requires 50% overlap. The fractional shift requires 75% overlap.

**Phase vocoder method for pitch scaling and its limitations**



For pitch shifting by a value X, all the frequencies must be multiplied by X. We use phase vocoder to that. We use a time scaling algorithm to stretch the sequence by a factor X. Then we sample the signal at X\*dt, where dt is the original sampling period. Thus, the rate at which music is playing remains the same but the frequency has been multiplied by X. The order in which we perform the time scaling and resampling affects the cost of the algorithm. It is wise to resample and then time scale for increasing the frequency as resampling gives a smaller signal. The opposite is true when decreasing the frequency

**Limitations**

The limitation of this technique is that the cost function is proportional to the frequency shift X. The cost becomes very high for large shifts in frequency.

We also see that at a time only one linear frequency shift can be achieved as all the frequencies get shifted by the same amount. Thus, adding harmonics, requires us to run the phase vocoder multiple times and increases cost

**Phase vocoder with peak-based pitch modulation**

**Procedure idea**

We identify the peaks in the short time fourier transform and then translate them to the desired frequencies. The blocks around the peaks are called ‘influence regions’ or bins. While shifting the peaks if we can maintain the amplitude and phases of the bins then the time domain transform corresponds to the same sinusoid at a different frequency. We use windows to break the signal into blocks.

Assume a signal given by



The stft of the signal is given by

  🡪 fourier transform function

If the frequency is shifted then the stft of the signal is given by



Note that when we shift the frequency from w to w+Δw , to keep the phase continuous we have to shift the peak phase by R\*Δw , where R is the hop size in the stft. Hop size is the number of samples between two blocks. The only information needed for this is the amount of shift Δw and the hop size R and thus, is easy to compute as compared to phase unwrapping.

Peak identification

We can use different criteria for identifying the peaks and setting the influence region. The simplest method is to use the amplitude of a bin. If the amplitude of a bin is greater than those of its two neighbors on either side then it can be classified as a peak.

We the find the influence region. This is set at the bin of lowest amplitude between two adjacent peaks.

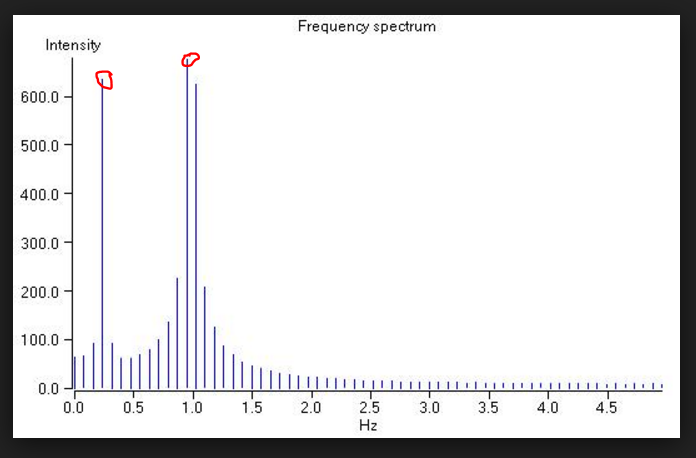


FIG. The two red dots indicate peaks as they are greater than the magnitude of neighboring two peaks on either side, the distance between adjacent impulses is called bin-distance

Frequency shift

The different effects can be simulated using various manipulations. Some examples are given below

Harmonize- we copy and shift all the peaks to locations representing the harmonizing factors

Chorus- we copy and shift all the peaks by small amounts close to the original locations.

Stretching-partial stretching involves applying quadratic correspondence in which the multiplying factor is the given frequency



Peak shifting

There are two cases depending on Δw. If Δw, is an integer multiple

Integer shift

If Δw, is an integer multiple of bin-distance, implying that there are integer number of bins between w and w+Δw then, shifting is merely copying values of original influence region to that around the shifted peak. When areas overlap, the magnitudes are added and if influence region is shifted into the negative axis, then we reflect it back with complex conjugation and add its values.

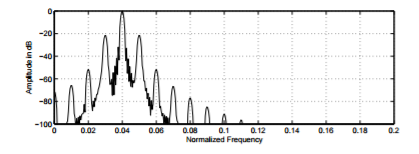
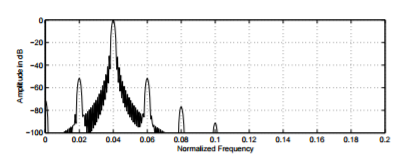


FIG. integer shift using 50% overlap hanning window

Fractional shift

If Δw is a fraction multiplied by bin-distance, then we have to interpolate as values are only known at integer multiples of bin distance. We can see that this is a fractional delay in the frequency spectrum.

For a 50% overlap side-bands of -21 db from peak amplitude are generated. While for the 75% overlap only -51 db side bands are generated. So the 75% overlap is required to minimize amplitude errors.



Phase alignment

To maintain phase continuity, we know that the phase must be multiplied by the quantity R\* Δw, where R is the hop size. This is done by multiplying the frequency bins in the influence region by the value



The computation cost is low as only calculating 1 sine, 1 cosine per peak and 1 multiplication per bin is required. The total shift in the frequency should be added together when calculating the shift from one peak to the next



Conclusion

We can see that there are a lot of advantages of this method as compared to phase vocoder

The cost of manipulation is independent of the modification and complex effects can be simulated easily which is not possible in phase vocoder.

Phase vocoder can only shift frequency in integer values whereas our algorithm can shift in integer as well as fractional values

The phase vocoder needs a 75% overlap whilst our algorithm can work with 50% for large sampled FFT. This reduces the cost and computation time

References

J. Laroche, “Time and pitch scale modification of audio signals,” in Applications of Digital Signal Processing to Audio and Acoustics, M. Kahrs and K. Brandenburg,

Audio effects theory, implementation and application ,Joshua Reiss and Andrew Mcpherson

] J. Laroche and M. Dolson, “Improved phase vocoder time-scale modification of audio,” to appear in May issue of IEEE trans. speech and audio proc., 1999

J. Laroche and M. Dolson, “Phase-vocoder: About this phasiness business,” in Proc. IEEE ASSP Workshop on app. of sig. proc. to audio and acous., New Paltz, NY, 1997