PHASE VOCODER

Radovan Kovačev
Master Degree Programme (2), FIT BUT
E-mail: xkovac27@stud.fit.vutbr.cz

Supervised by: David Bařina
E-mail: ibarina@fit.vutbr.cz

Abstract: The aim of this paper is to give reader a brief explanation of what phase vocoder is and how it works, in uncomplicated way. Firstly, article deals with concept of time-frequency analysis. Then, the featured algorithm is illustrated. Finally, achievements of implementation are presented and all is summed up.

Keywords: time-frequency analysis, short time Fourier transform, phase vocoder

1 PREFACE

In the area of computer multimedia production, there is often a need for signal synchronization. Modern music creation approaches involve variety of manipulations with captured sound samples, including speeding them up or slowing down, while preserving the pitch. Otherwhile some transposition is required. Besides these effects, phase vocoder can also be good for vocal correction as in popular commercial VST plugin called Auto-Tune, which is based on it.

2 TIME-FREQUENCY ANALYSIS

The point of time-frequency analysis (TFA) is to express spectral evolution of non-stationary signal within time-frequency plane. The major tool of TFA is short-time Fourier transform (STFT), expressly defined as:

$$F(\tau, f) = \int_{-\infty}^{\infty} x(t)w(t-\tau)\mathrm{e}^{-\mathrm{i}2\pi ft}\mathrm{d}t \tag{1}$$

where w represents window function, which length affects temporal and spectral localization. According to Heisenberg uncertainty principle, resolution of time-frequency plane is limited by grid. Higher time accuracy is at the expense of frequency precision and conversely. These features are mutually exclusive and that can be seen in Figure 1.

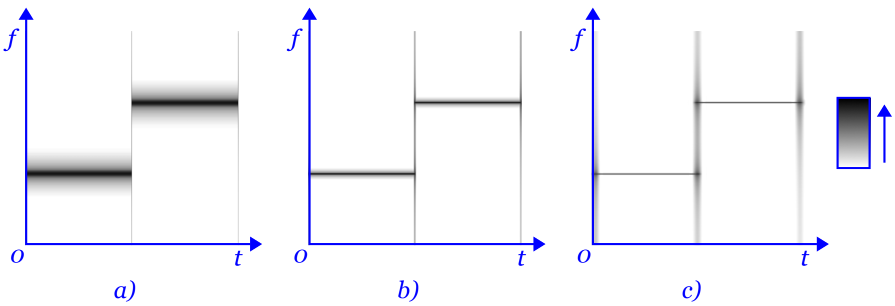

Figure 1: Progression of two sine waves.

Picture shows three STFT spectrograms computed using short (a), moderate (b) and long window (c), so different accuracies in temporal and spectral domain are obvious.

Lately, in attempt of adapting TFA to common signals, a new decomposition method called wavelet transform was developed. It's multiresolutional and quite improves in perfomance over STFT. This trasform is given by equation:

$$W(\tau, a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-\tau}{a}\right) \mathrm{d}t \tag{2}$$

where ψ denotes wavelet, a localized function that meets special criteria.

Despite both techniques bring high redundancy, they are still very useful and widely applied to practice. For example in climate research, heart monitoring, seismic signal denoising, characterization of turbulent intermittency, image compression, studies of financial indices and so on [1].

3 PRINCIPLE

The phase vocoder stands for STFT based time-frequency analysis/processing/resynthesis system. In the strict sense, it's an audio processor capable of altering duration of sound with no change of it's spectral properties and vice versa. Phase vocoder was invented by Flanagan and Golden of Bell Telephone Laboratories, who developed the first program in 1966. Originally, it was intended to be a coding method for reducing the bandwidth of speech signals [3].

The algorithm is quite simple to understand and goes like this. Signal data coming from input stream are read by frames that overlap. Each of them is modulated with envelope to get better frequency response. Windowed chunk is then spectral analyzed. Phase for every frequency is modified to keep coherence after synthesis. For pitch-shifting, further frame length change using interpolation becomes necessary. At the end, the processed blocks are windowed one more time and written to output stream with new overlap factor.

The key step is shown in Figure 2. As it can be seen, the frame distance is at first adjusted from a to b, so there develops a discontinuity and the phase must be reccurently recalculated. To make things simple, segments are not modulated and consist of only one spectral component.

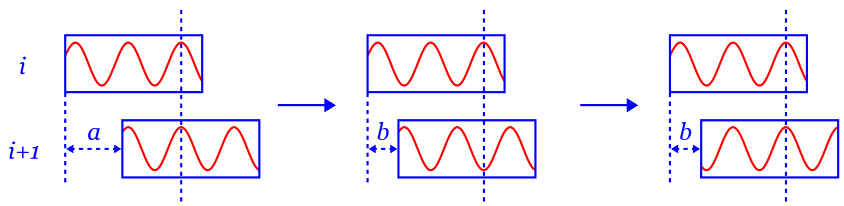

Figure 2: Length compression.

4 IMPLEMENTATION

The application code is written in pure C with no use of non-standard libraries. Program allows user to perform time-stretching and pitch-shifting to any extent. In addition, it plots spectrograms.

Core algorithm utilizes the most essential $N \log N$ FFT [2], optimized for real signals and therefore brings low latency. File operations work pseuoparallel, so real-time effect would be possible to recreate from this with a little effort, even for DSP unit. Overall computional complexity is linear (LTI system) with constant (static) memory allocation.

5 RESULTS

Experimenting with phase vocoder, it should be noticed, that squeezed signal sounds more natural than extended. This observation can have logical explanation in decreasing and respectively increasing data density. Expanded version just lacks enough information. This wouldn't happen with analog.

Visual demonstration of phase vocoding captures Figure 3, where a) is input b) shortened c) lengthened d) lowered and e) raised signal. The sample is so trivial to perceive changes clearly. In case of d) and e), some interpolation artifacts can be seen and that's because of logarithmic magnitude scale.

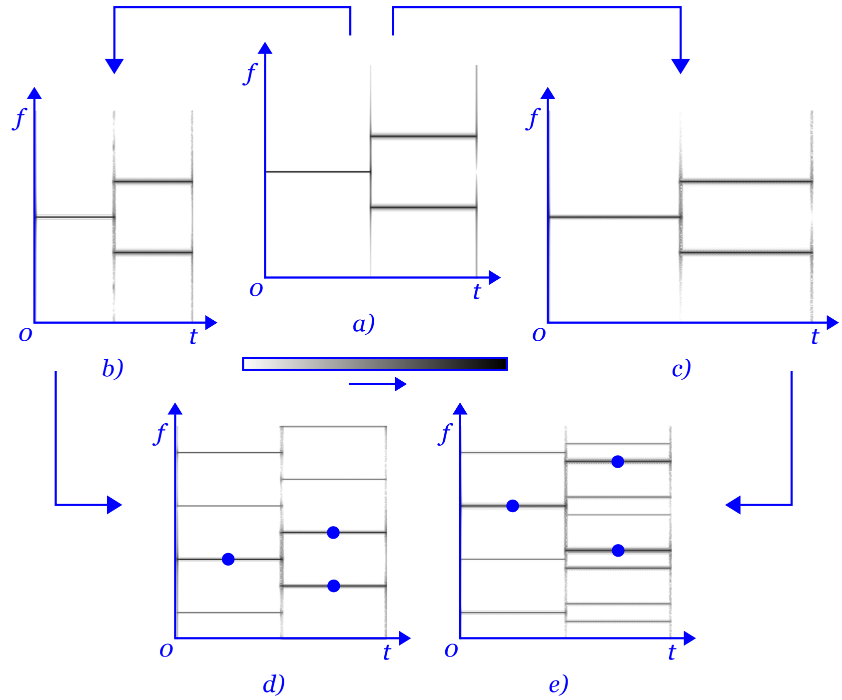

Figure 3: System operation.

6 CONCLUSION

The idea of TFA and phase vocoder evidently make sense, which was proved by implementation, that operates properly. Future ehnancement may be to add an instantaneous frequency detection for automatic pitch-shifting or tracking the spectral peaks for additive synthesis with sinusoidal oscillators.

7 ACKNOWLEDGEMENT

This work has been partially supported by the Research Plan MSM0021630528, the grant BUT FIT-S-11-2 and IT4Innovations Centre of Excellence project CZ.1.05/1.1.00/02.0070.

REFERENCES

[1] Mallat, S.: *A Wavelet Tour of Signal Processing.* Academic Press, 2008, ISBN 0123743702

[2] Proakis, J.G.; Manolakis, D.G.: *Digital Signal Processing.* Pearson, 2006, ISBN 0131873741

[3] Roads, C.: *The Computer Music Tutorial.* The MIT Press, 1996, ISBN 0262680823