

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

ÚSTAV TELEKOMUNIKACÍ

Ing. Michal Trzos

MODERN METHODS OF TIME-FREQUENCY WARPING OF SOUND SIGNALS

MODERNÍ METODY BORCENÍ ČASOVÉ A KMITOČTOVÉ
OSY ZVUKOVÝCH SIGNÁLŮ

ZKRÁCENÁ VERZE PH.D. THESIS

Obor: Teleinformatika

Školitel: Ing. Jiří Schimmel, PhD.

Oponenti: ??
??

Datum obhajoby: x.5.2015

KLÍČOVÁ SLOVA

Harmonická transformace, Fan-Chirp transformace, FChT, QFFT, PTDFT

KEYWORDS

Harmonic Transform, Fan-Chirp Transform, FChT, QFFT, PTDFT

Disertační práce je k dispozici na Vědeckém oddělení děkanátu FEKT
VUT v Brně, Technická 10, Brno, 616 00

© Trzos Michal, 2015

ISBN 80-214-

ISSN 1213-4198

CONTENTS

Introduction	4
1 Thesis Objectives	5
2 State of the Art	6
2.1 Harmonic Transform	6
2.2 Fan-Chirp Transform	7
3 Research results	8
3.1 Reducing The Number Of Computations Of The Harmonic Transform	8
3.2 Fast Harmonic Transform	9
3.2.1 Inverse Fast Harmonic Transform	10
3.3 Estimation of Fundamental Frequency Change Using Gathered Log-Spectrum	10
3.4 Estimation of Fundamental Frequency Change Using Analysis-by-Synthesis Approach	13
3.5 Computational Load	15
3.6 Effect of Aliasing	16
3.7 Experiments	17
4 Conclusion	21

INTRODUCTION

From all mechanisms of communication, sound communication is by far the most widely used by humans and at the same time easily processed using modern technology, namely digital signal processing. Most mammals, including humans, communicate using air stream modulation ranging in frequency from infrasound (whales) to ultrasound (bats). If the air stream modulation is constant, the produced sound can be approximated using an impulse train. In the frequency domain, the impulse train consists of a fundamental frequency and harmonics at integer multiplies of the fundamental frequency. Such signal can therefore be effectively analyzed using traditional tools like the Fourier Transform. However if the air stream modulation changes in time, as is the case of most real signals, its frequency components also change in time. While frequency variance of the fundamental frequency may not be significant, it multiplies for each additional harmonic contained in the signal. When using Fourier Transform to analyze such signal, the higher harmonics may span over several frequency bins of the analyzed time interval depreciating the accuracy of harmonic parameters that can be acquired from the signal.

There are many applications that rely on the analysis of harmonic signals with time-varying components. Most of them deal with speech signals for speech coding, gender and age classification, detection of alcohol intoxication, emotion detection, or jitter estimation in Parkinsonian speech. Some musical instruments can be played in a way that causes fundamental frequency modulation like viola, violin, trombone, or guitar while some instruments create frequency modulation by their nature like the Theremin or the Leslie speaker. Also most synthesizers can be modulated using the pitch wheel which enables continuous variation of the fundamental frequency. Analysing such signals may be performed with higher precision with a method that enables to take time-variant fundamental frequency into account.

This thesis therefore focuses on the representation of non-stationary signals with time-varying components. First it provides a summary of the state-of-the-art methods with main focus on Fan-Chirp Transform and Harmonic Transform. Then the focus turns solely on the Harmonic Transform and its computational demands which prevent its efficient use. A prerequisite to computing Harmonic Transform is knowledge of fundamental frequency change and an approach to decrease its estimation is presented. However the goal is decrease in computational complexity, which is presented as the Fast Harmonic Transform. This introduces some artifacts to the signal which is covered in the text. Then two algorithms for fundamental frequency estimation are presented. One is based on the gathered log-spectrum and the other on analysis-by-synthesis approach. Both algorithms are applied to a speech signal to compare their output. The thesis finishes with experiments on real signals.

1 THESIS OBJECTIVES

From the methods for representation of non-stationary signals with time-variant frequency components presented in the previous chapter this thesis will deal with the Harmonic Transform. Specifically with decreasing its computational demands. Knowledge of fundamental frequency change is required before computing the Harmonic Transform. This is done by Spectral Flatness Measure and our first focus will be on optimizing its computation. Unfortunately, the Harmonic Transform computation still employs $\mathcal{O}(N^2)$ computational complexity. So the next focus will be on obtaining a Harmonic Transform which employs subquadratic computational complexity. This will be attempted substituting the time-warping kernel of Harmonic Transform with time-warping of the time axis. Since we usually only have discrete signals available, it is necessary to use interpolation which introduces noise into the signal. This renders Spectral Flatness Measure ineffective for the computation of fundamental frequency change as will be shown in the next chapter. Therefore a different method of fundamental frequency change estimation is needed. Two methods will be presented in this thesis. The first method computes fundamental frequency change using gathered log-spectrum which performs gathering of the logarithm of the magnitude spectrum at the places of the fundamental frequency and its multiples. The second method selects the optimal fit of fundamental frequency change by comparing the reconstruction error of the harmonic part of the signal which is estimated using the Harmonic Transform centered on the fundamental frequency. Both of these methods will be tested on the same speech signal to compare their approach.

Since the Fast Harmonic Transform uses interpolation for its fast computation, there will inevitably be artifacts caused by the interpolation. This will be even more pronounced in the signal reconstructed using Inverse Fast Harmonic Transform from the harmonic domain. The reconstruction error will be measured for several interpolation methods. Another artifact present in the Fast Harmonic Transform image is aliasing and it will be addressed using oversampling and evaluated for different oversampling factors and interpolation methods.

To summarize the goals of this thesis, they can be divided into these main areas:

- Fast Harmonic Transform Algorithm
- Fast Inverse Harmonic Transform Algorithm
- Computational load of the Fast Harmonic Transform
- Fundamental frequency change estimation using gathered log-spectrum
- Fundamental frequency change estimation using analysis-by-synthesis approach
- Aliasing artifacts and anti-aliasing by oversampling
- Experiments on real frequency-modulated signals

2 STATE OF THE ART

Many harmonic signals, including speech and music, exhibit frequency modulation caused by varying fundamental frequency. A traditional instrument for the analysis of speech and musical signals is Fourier Transform (FT). The ability of the FT to represent frequency content of a signal diminishes if the signals contains components with varying frequency [1, 2]. One solution of this problem is to use Warped Fourier Transform (WFT) [3], where the signal is frequency or time warped [4] before applying the FT, giving birth to warped wavelets [5, 6]. This operation can be interpreted as change of the signal's scale for the conversion of time-varying frequency components to frequency invariant components. The scaling operation can be generalized using the Scale Transform [7–9], where the scale is taken as a physical property of the signal, or the scaling operation can be integrated into the definition of transformation, as in Harmonic Transform [10]. Speech signals and other harmonic signals with a formant structure require a method to preserve the formant structure if modified. This can be done efficiently using frequency warping [11, 12].

There are other means of representation of signals with variable frequency components which are based on several models of speech. A family of transforms is based on the similarity of voiced speech to a chirp-periodic signal. Fan-Chirp Transform [13, 14] is suitable for signals with frequency components varying linearly on fan geometry, a property providing it with the best representation of chirp-like signals.

2.1 Harmonic Transform

Harmonic transform has been introduced in [10] and it is based on [15] [16]. Its main difference from Fourier transform is the integrated time-warping function. It is defined as

$$S_{\phi_u(t)}(\omega) = \int_{-\infty}^{+\infty} s(t) \phi'_u(t) e^{-j\omega \phi_u(t)} dt, \quad (2.1)$$

where $\phi_u(t)$ is a unit phase function, which is the phase of the fundamental harmonic component divided by its nominal instantaneous frequency [10], and $\phi'_u(t)$ is first derivation of $\phi_u(t)$. The $\phi_u(t)$ is required to be invertible and differentiable on $(-\infty, +\infty)$. When the $\phi_u(t) = t$, the HT reverts to the FT. The inverse harmonic transform (IHT) is defined as [10]

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{\phi_u(t)}(\omega) e^{j\omega \phi_u(t)} d\omega. \quad (2.2)$$

The DHT variant aligned with the fundamental frequency is defined as [17]

$$S(k) = \sum_{n=0}^{N-1} s(n) \alpha'(n) e^{-j \frac{2\pi k f_r}{f_s} \alpha(n)}, \quad (2.3)$$

where f_r is the fundamental frequency and $k = 1, \dots, K$ is the number of harmonics.

2.2 Fan-Chirp Transform

For a signal $x(n)$, which is a discrete-time version of the signal $x(t)$ at sampling frequency f_s , the discrete-time FChT is defined as [14]

$$X(f, \alpha) = \sum_{n=0}^{N-1} x(n) \sqrt{\phi'_{\hat{\alpha}}(n)} e^{-j 2\pi \frac{k}{N} \phi_{\hat{\alpha}}(n)}, \quad (2.4)$$

where k is the frequency bin index, N is the number of segment samples, $\hat{\alpha}$ is related to its continuous-time counterpart $\hat{\alpha} = \alpha/f_s$, and $\phi_{\hat{\alpha}}$ is the following mapping, bijective in $[0, N]$

$$\phi_{\hat{\alpha}} = \left(1 + \frac{1}{2} \hat{\alpha} (n - N) \right) n. \quad (2.5)$$

While the discrete-time FChT can be computed directly using (2.4), computational load of the direct version is quadratic. A fast version of the FChT operates reformulating the FChT as the FFT of a time-warped signal, substituting $\tau = \phi_{\alpha}(t)$ thus significantly reducing computation [13]. The FChT with the variable substitution becomes [18]

$$X(f, \alpha) = \int_{\phi_{\alpha}(-\frac{T}{2})}^{\phi_{\alpha}(\frac{T}{2})} \tilde{x}(\tau) \tilde{\rho}(\tau) e^{-j 2\pi f \tau} d\tau, \quad (2.6)$$

where $\tilde{x}(\tau)$ is a time-warped version of the signal $x(t)$ and $\tilde{\rho}(\tau)$ is a scaling function on the time-warped axis. In discrete time equation (2.6) can be written as

$$X(k, \hat{\alpha}) = \sum_m \tilde{x}(m) \tilde{\rho}(m) e^{-j 2\pi \frac{k}{K} m}, \quad (2.7)$$

where the range of m in (2.7) is derived from the relationship $\phi_{\alpha}(-\frac{T}{2}) \geq \tau \geq \phi_{\alpha}(\frac{T}{2})$, which yields

$$M \left(\frac{1}{8} \hat{\alpha} N - \frac{1}{2} \right) \leq m \leq M \left(\frac{1}{8} \hat{\alpha} N + \frac{1}{2} \right). \quad (2.8)$$

3 RESEARCH RESULTS

This chapter deals with efficient implementation of the Harmonic Transform. The original implementation requires $\mathcal{O}(N^2)$ operations. The first approach to reduce the number of computations required to compute HT is to reduce the number of operations for computation of spectral frequency measure. This is done by exploiting redundancy in its algorithm.

This however still leaves an algorithm with quadratic computational complexity, so the research is then focused on producing an algorithm with subquadratic computational complexity. This is achieved by time-warping the input signal, where the relationship between the warped axis and original axis is given by the transformation kernel of the HT.

3.1 Reducing The Number Of Computations Of The Harmonic Transform

One of the crucial steps in computation of the Harmonic Transform is to estimate the fundamental frequency change of the analyzed signal. So far, algorithm based on SFM has been used. When exploring this algorithm, several observations have been made. The SFM has several minimums and if a search algorithm was used, it could fall into local minimum. It is also noteworthy that it is possible the harmonic transform $|\text{DHT}(a, k)|$ will be equal to zero for some values of k , which could mean that the spectral flatness will be zero for all a . Removing zero values solves this problem and leads to band-limited spectral flatness measure [19].

Harmonic spectrum of the Harmonic Transform is not complex conjugated even for real signals (which is true for Fourier transform). From the frequency axis point of view, the unit phase function $\phi_u(t)$ shifts the spectrum towards lower frequencies if a is positive, and to higher frequencies if it is negative. For harmonic signal analysis, only left side of the spectrum is useful, because it appropriately represents non-stationary harmonic signal. Using the modified spectral flatness measure (MSFM) [19]

$$\arg \min_a \text{MSFM}(a) = \frac{\sqrt{\prod_{k=0}^{N/2} |\text{DHT}(a, k)|}}{\frac{1}{N/2+1} \sum_{k=0}^{N/2} |\text{DHT}(a, k)|} \quad (3.1)$$

we can get function of a which has clearly defined minimum [20]. This is caused by using only left side of the spectrum when computing SFM and it consequently leads to reducing the number of operations needed to compute spectral flatness by $\frac{N}{2} - 1$ [20].

3.2 Fast Harmonic Transform

The number of operations in direct computation of the HT raises quadratically, similarly to direct computation of Fourier transform. The goal of this section is to present an algorithm to compute the HT which shows sub-quadratic complexity. When there is a transform with quadratic complexity, then its sub-quadratic form is referred to as the fast version of the transform. In this case it is the Fast Harmonic Transform (FHT).

Discrete-Time Fast Harmonic Transform

The DFHT can be written as

$$S(k, a) = \sum_{n=0}^N \tilde{s}(n) \tilde{\rho}(n) e^{-j2\pi \frac{k}{K} n} \quad (3.2)$$

which is a FFT of the product $\tilde{s}(n)\tilde{\rho}(n)$ which is the uniformly sampled product $\tilde{s}(\tau)\tilde{\rho}(\tau)$. Since we usually only have discrete signals available, we will use discrete-time intervals n even though its value can be non-integer. Any values at non-integer intervals will be enumerated using interpolation from the signal samples. Now to get a discrete-time counterpart we take $\alpha_a(n)$ which is a quadratic function and its inverse $\alpha_a^{-1}(n)$ yields two results. The result of interest is

$$\psi_a(n) = \frac{N}{2} - \frac{N}{a} + \frac{N\sqrt{(\frac{a^2}{4} - a + \frac{2an}{N} + 1)}}{a}, \quad (3.3)$$

where n is sample index and N is number of samples [21].

With (3.3) we can define the discrete-time time-warped signal as

$$s_a(n) = \tilde{\rho}(n) \tilde{s}(\psi_a(n)), \quad (3.4)$$

where $\tilde{\rho}(n) = \phi'_a(\psi_a(n))^{-1}$ is the scaling factor which can be written as

$$\tilde{\rho}(n) = \left(-\frac{a}{2} + \frac{\frac{N}{2} - \frac{N}{a} + \frac{N\sqrt{a^2/4 - a + \frac{2an}{N} + 1}}{a}}{N} + 1 \right)^{-1} \quad (3.5)$$

and $\tilde{s}(\psi_a(n))$ is the time-warped signal [21]. The last step to compute the HT is using FFT on the time-warped signal $s_a(n)$ as follows [21]

$$S(k, a) = \sum_{n=0}^{N-1} s_a(n) e^{-j2\pi \frac{k}{N} n}. \quad (3.6)$$

Now we have a Fast Harmonic Transform for harmonic signals with linear frequency change which in the next step will be turned into an algorithm which will enable

Tab. 3.1: SNR (dB) of a speech signal *micf01sa02* reconstructed using IFHT from a harmonic spectrum obtained by FHT.

interpolation method	oversampling		
	1x	2x	4x
linear	17.9	28.0	37.1
cubic	22.0	37.7	42.6
spline	28.0	42.4	42.9

its use for analysis and synthesis in the harmonic domain. Fast implementation of the Harmonic Transformation is based on (3.6), though its actual implementation employs several improvements. Block diagram of the Harmonic Transform algorithm is shown in Fig. 3.1

3.2.1 Inverse Fast Harmonic Transform

Inverse Fast Harmonic Transform (IFHT) is the inverse transform to the Fast Harmonic Transform. It can be used to obtain a time domain signal from a harmonic spectrum and its estimated fundamental frequency slope a . The IFHT is defined as

$$s(n) = \frac{1}{N} \sum_{k=0}^{N-1} S(k, a) e^{j2\pi \frac{k}{N} n}. \quad (3.7)$$

An algorithm to compute the IFHT is very similar to the algorithm of FHT with reversed block order. The block diagram is in Fig. 3.2. Description of the blocks follows.

3.3 Estimation of Fundamental Frequency Change Using Gathered Log-Spectrum

A block diagram of this method can be seen in Fig. 3.3. Its principle is computation of gathered log-spectrum for a predefined range of fundamental frequencies and fundamental frequency changes based on the nature of the analyzed signal. A (a, f_0) plane is constructed from the gathered log-spectrum values which represent pitch salience and the most likely candidates for fundamental frequency are represented as peak values. For signals with dominant first harmonic component the first candidate with highest value is usually equal to the fundamental frequency in the analyzed signal. The resulting fundamental frequency f_0 and its slope a is taken from the maximum value of the gathered log-spectrum.

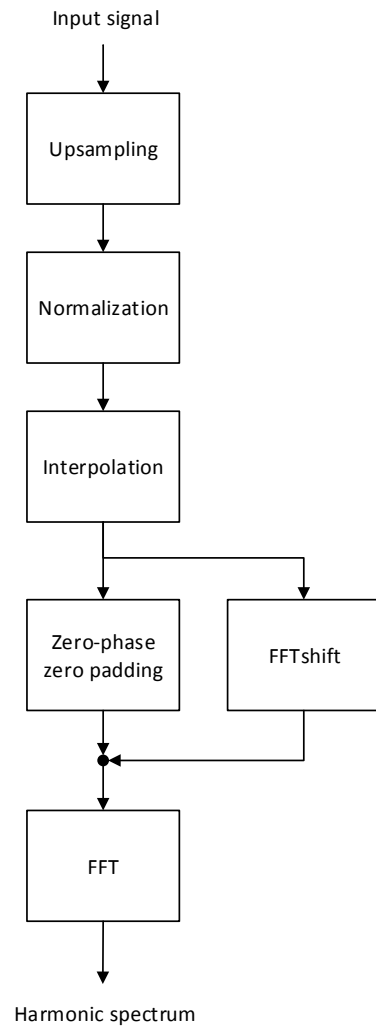


Fig. 3.1: Block diagram of the forward Fast Harmonic Transform.

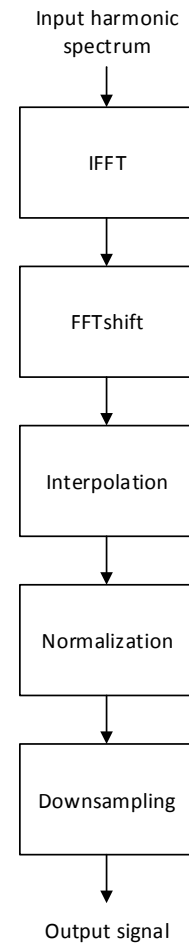


Fig. 3.2: Block diagram of the Inverse Fast Harmonic Transform computation.

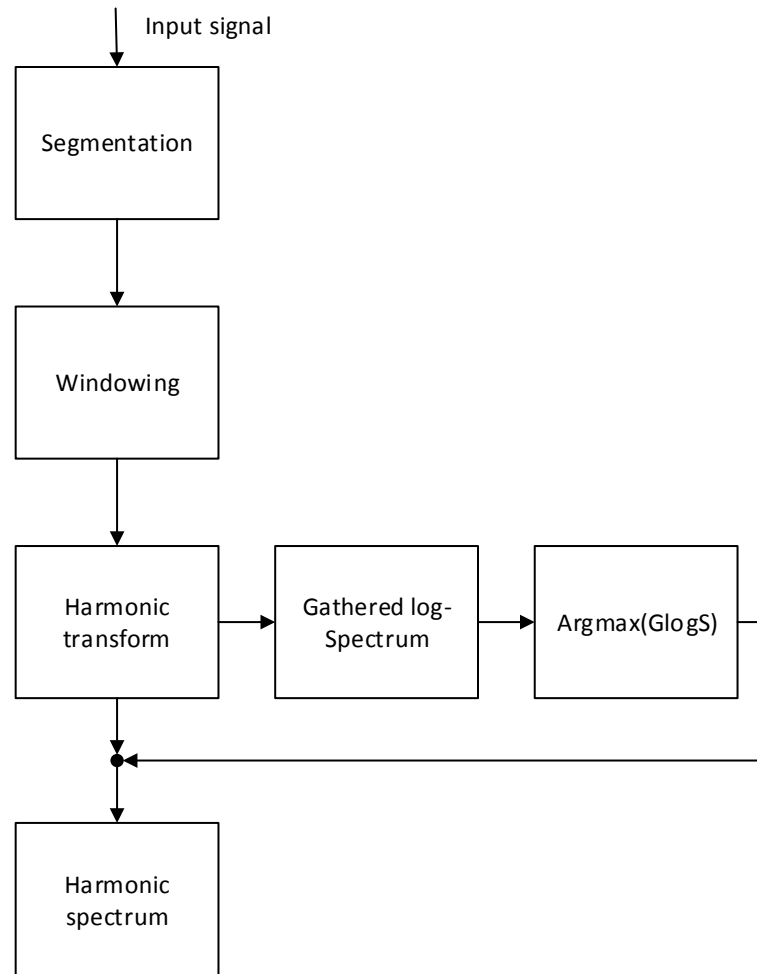


Fig. 3.3: Block diagram of Harmonic Transform computation with f_0 estimation using gathered log-spectrum.

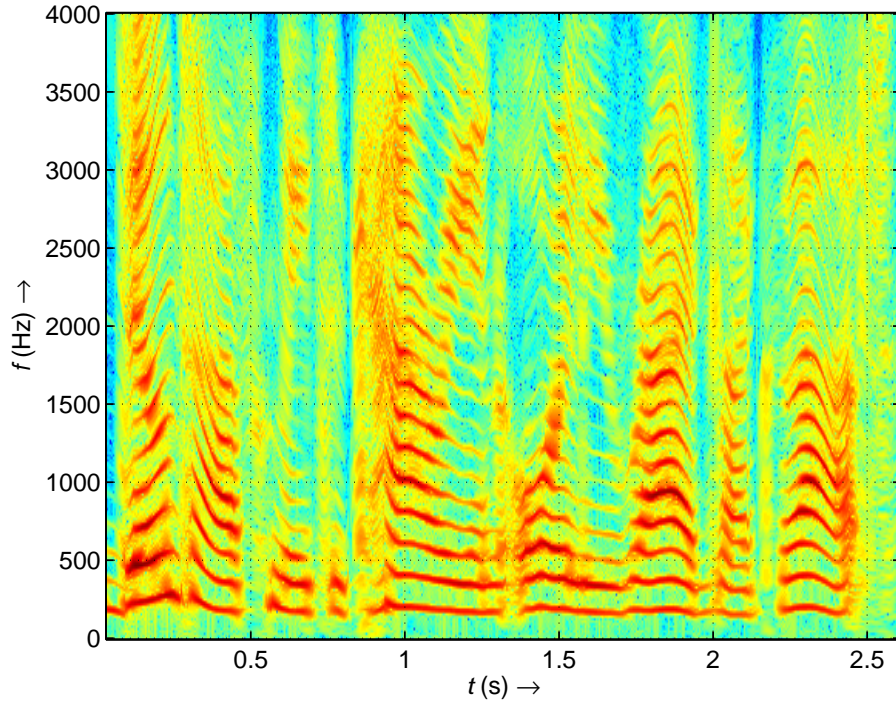


Fig. 3.4: Spectrogram of the *micf01sa02* signal obtained using Fast Harmonic Transform with gathered log-spectrum as the f_0 change estimation algorithm.

To show a typical output of the presented algorithm, it has been run on an speech signal *micf01sa02* which is an utterance “Don't ask me to carry an oily rag like that” pronounced by a female speaker with parameters $M = 511$, $NFFT = 511$, overlap = 5 ms, $f_s = 8$ kHz, $n_H = 4$, for $f_0 \in < 80; 350 >$ and $a \in < -0.3; 0.3 >$ without oversampling. Spectrogram constructed from the outputs of Harmonic Transform is shown in Fig. 3.4 and a STFT spectrogram is shown in Fig. 3.5 for reference. It is evident the Harmonic Transform based spectrogram has sharper peaks without spectral smearing where a harmonic structure is present in the signal, specifically in the higher frequencies.

3.4 Estimation of Fundamental Frequency Change Using Analysis-by-Synthesis Approach

In this approach we will use the (a, f_0) plane to estimate the fundamental frequency as in 3.3 but with harmonic-to-noise ratio instead of pitch salinity. This approach assumes analysis of signals which are composed of a fundamental frequency and its harmonics. We will try to estimate harmonic parameters of each harmonic of such signal where hypothetical number of harmonics n_H , range of fundamental frequencies

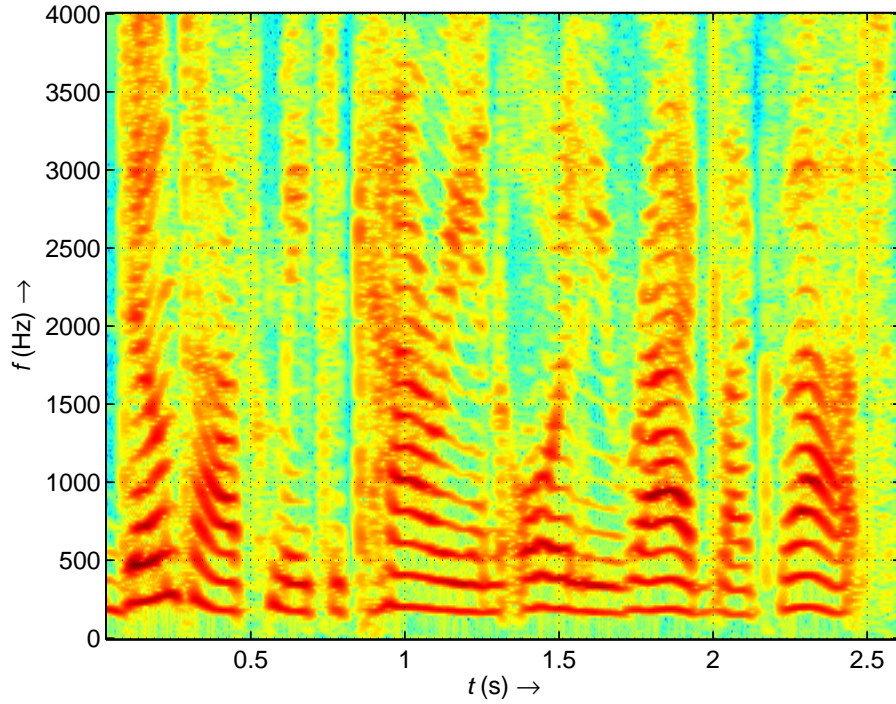


Fig. 3.5: Spectrogram of the signal *micf01sa02* obtained using STFT.

f_0 and range of fundamental frequency changes a is based on previous knowledge of the nature of the analyzed signal. After the harmonic parameters have been estimated, they are used to construct the harmonic part of the analyzed signal which is then subtracted from the analyzed signal to get the residual signal. Then harmonic-to-noise ratio is computed from the harmonic and residual signal for all values of a and f_0 which are then assembled on the (a, f_0) plane. For a and f_0 that match the analyzed signal there will be a peak in the (a, f_0) plane and these values are evaluated as the final values.

FFT cannot be used to compute (2.3) though its computational complexity is $\mathcal{O}(kN)$, where k is the number of harmonic components and N is length of the transformation. Computational requirements can be kept reasonable through suitable choice of input parameters. Block diagram of the algorithm can be seen in Fig. 3.6.

The algorithm has been tested on a signal *micf01sa02* with the same parameters as in case of the method presented in Section 3.3: $M = 511$, $NFFT = 511$, overlap = 5 ms, $f_s = 8$ kHz, $n_H = 4$, for $f_0 \in < 80; 350 >$ and $a \in < -0.3; 0.3 >$ without oversampling. From Fig. 3.7 we can see the harmonic spectrogram provides much sharper peaks compared to the STFT spectrogram in Fig. 3.5 and it is very similar to the harmonic spectrogram obtained using gathered log-spectrum as can be seen in Fig. 3.4. There are also parts where the harmonic spectrogram provides doubtful

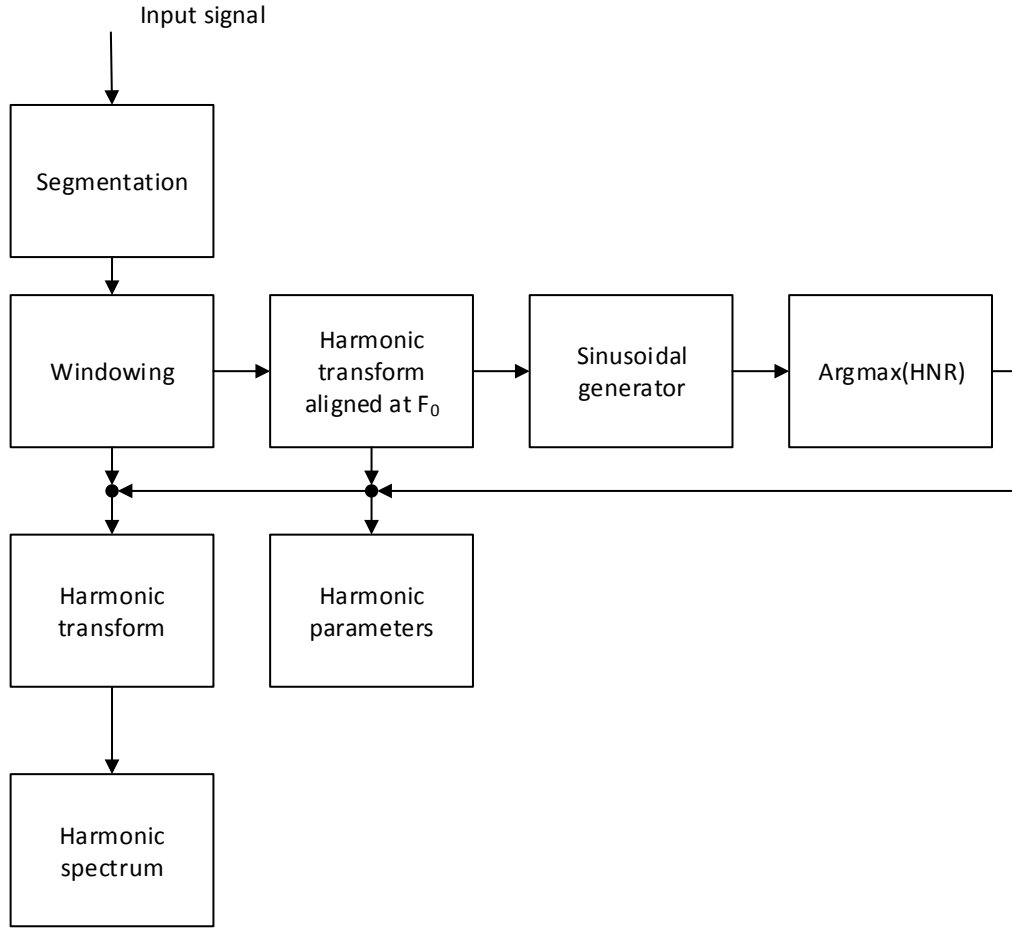


Fig. 3.6: Block diagram of Fast Harmonic Transform algorithm using harmonic parameters for f_0 change estimation.

results occurring usually at transients e.g. at time intervals (0.8 s;1 s) and (2.3 s; 2.5 s).

3.5 Computational Load

The computational load of the fast algorithm can be enumerated using the number of operations required for analysis of one segment of length N . In normalization stage, the input signal $x(n)$ is multiplied by the window function which has been divided by the scaling factor $\phi'(n)$. Warped index computation estimates time instants of the signal time-warped according to the warping function $\psi_a(n)$. The time-warped discrete-time signal $s_a(n)$ is obtained using interpolation from the normalized input signal $z_a(n)$. Finally the output harmonic spectrum $S(k, a)$ is computed using FFT.

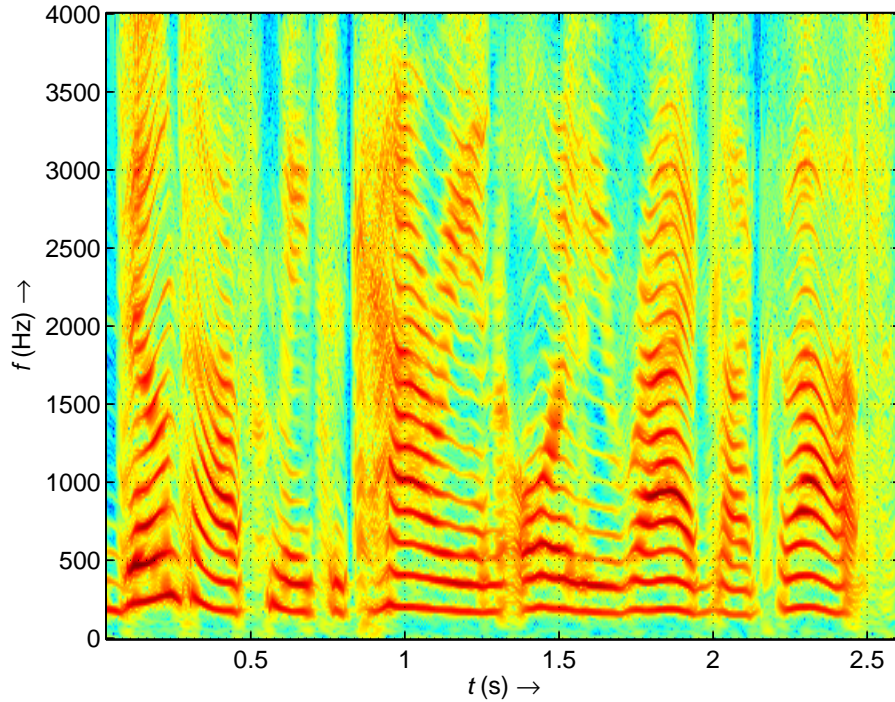


Fig. 3.7: Spectrogram of the signal *micf01sa02* obtained using the analysis-by-synthesis method.

The resulting computational load is $N(\log N + 7)$ for the Hermite spline interpolation and $N(\log N + 5)$ for the linear interpolation.

3.6 Effect of Aliasing

The Fast Harmonic Transform uses interpolation of the input signal which introduces errors, namely, aliasing. To demonstrate the effect of aliasing we have used *test signal* which is a linear chirp with 17 harmonics. Time warping performed using (3.3) maps one axis with equidistant intervals to a time-warped axis where the intervals between samples get shorter towards one of the ends of analysis segment. This causes the signal on the warped axis to be undersampled. Aliasing can be seen as a noise floor which increases with frequency.

One of the straightforward means of diminishing aliasing is oversampling. Oversampling consists of increasing the sampling frequency by adding zeroes to the signal and then filtering the signal by a low-pass filter to eliminate mirroring artifacts. The resulting signal will have a multiple number of samples which in principle reduces the intervals between samples of the signal on the original axis and therefore the time-warped signal is interpolated with higher precision. This also allows us to use a cheaper interpolation method, if advantageous. A case where linear interpolation

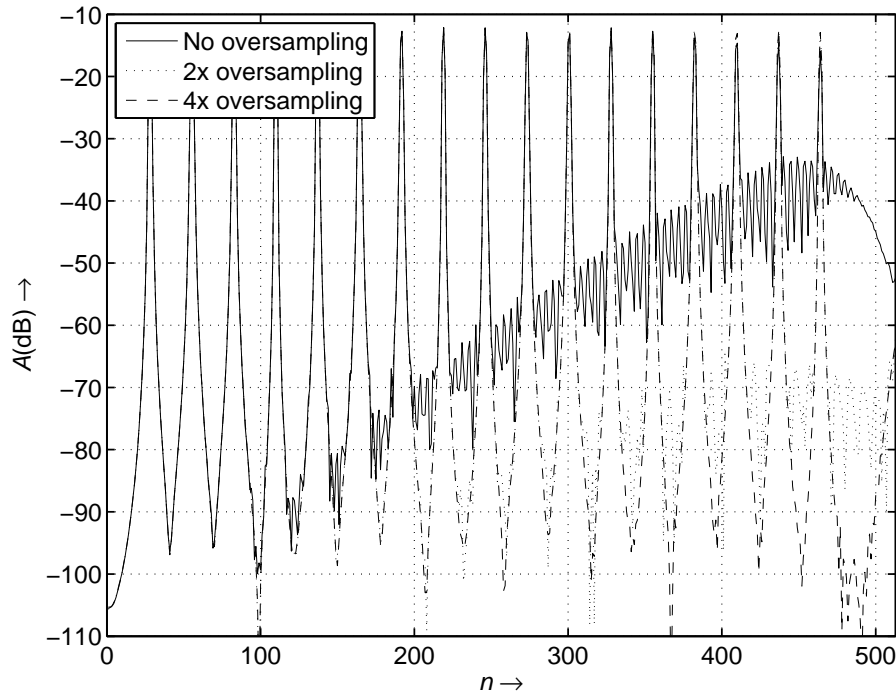


Fig. 3.8: The effect of oversampling on aliasing. Fast Harmonic Transform with linear interpolation was used on *test signal*.

was used on the *test signal* with 2x and 4x oversampling is shown in Fig. 3.8. A more thorough analysis has been performed on signal *micf01sa02* as shown in Tab. 3.1.

3.7 Experiments

Since we are analyzing real signals, there is no ground truth for the signal's harmonic parameters at each instant as opposed to analyzing synthesized signals, where the parameters are known and can be directly compared. Therefore we will analyze the signal using the ABS algorithm and use it to extract the fundamental frequency which will be used as input to harmonic parameter estimation. The signal will then be reconstructed using the harmonic parameters when using the knowledge of fundamental frequency slope and without this knowledge. This will produce a synthetic harmonic signal, an estimate of the input signal, with (further referred to as ABS-FM) and without frequency modulation (ABS-S). The ABS-S algorithm is essentially the same algorithm as ABS-FM with $a = 0$. This synthetic harmonic signal will then be subtracted from the input signal, leaving a residual signal. The better the harmonic parameter estimation, the lesser the residual signal energy. By measuring the harmonic-to-noise ratio for different signals with frequency modulation while using the knowledge of fundamental frequency change and without it,

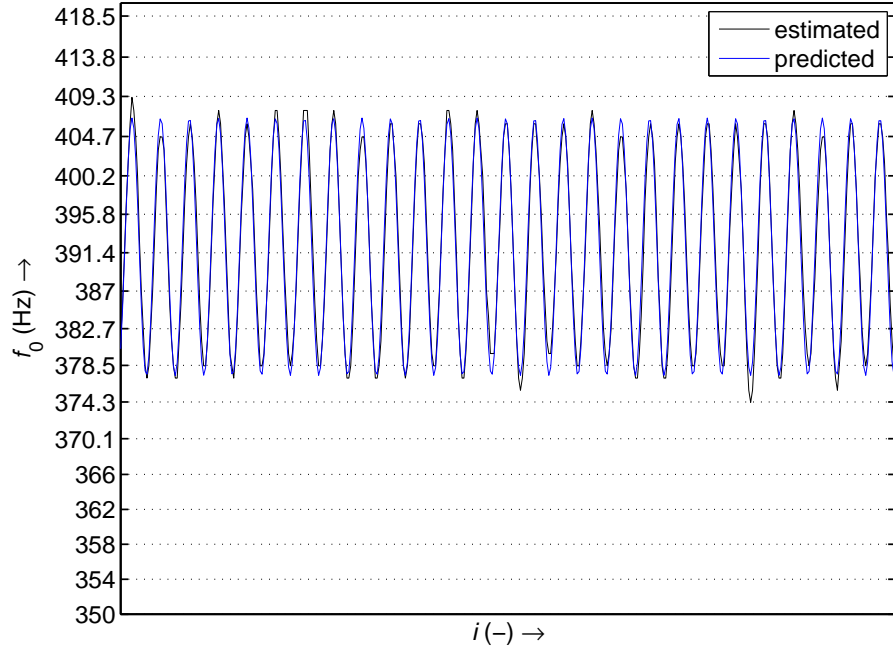


Fig. 3.9: Fundamental frequency of vocal sample *salvation* with artificial vibrato.

we can quantify the increase of harmonic parameter estimation accuracy which we get by using ABS-FM algorithm. So far, the Harmonic Transform has been used on speech signals which are usually conveniently sampled at 8 kHz. Yet, for many applications higher sampling frequencies are required. Experiments in this section are done on audio signals with sampling frequency 44.1 kHz.

Artificial vibrato

In this experiment we would like to apply frequency modulation on a harmonic signal with known and nearly stationary fundamental frequency to compare the ability to estimate harmonic parameters from a signal in our system from section 3.4 when using ABS-FM and ABS-S. The selected harmonic signal is a decaying vocal excerpt with nearly stationary fundamental frequency. The frequency modulation is created using a vibrato audio effect. This can also be observed from Fig. 3.9 which shows the computed vibrato as the predicted sinusoid and the estimated sinusoid shows the estimated fundamental frequency from the ABS-FM algorithm. The difference between ABS-FM and ABS-S as shown in Fig. 3.10 shows increase in harmonic component separation at intervals with frequency modulation.

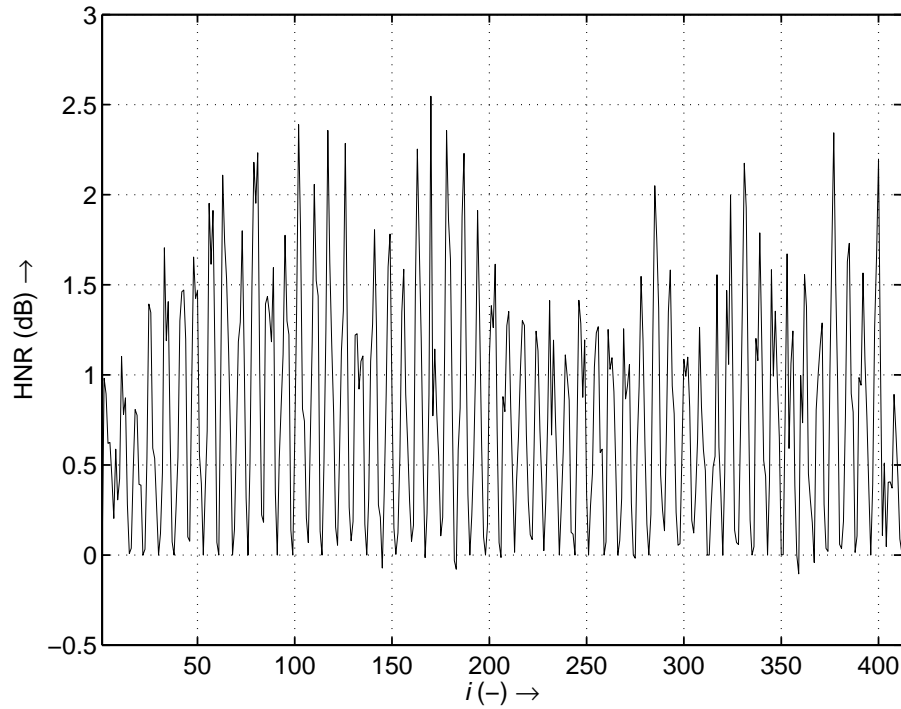


Fig. 3.10: HNR increase of ABS-FM over ABS-S of reconstructed harmonic part of the sound sample *salvation*.

Viola

This experiment has been performed on a *viola* sound sample. It contains glissando and vibrato, which are both frequency modulation techniques on stringed instruments. Segments at which the techniques are used can be seen from the fundamental frequency estimated by the ABS-FM algorithm in Fig. 3.11. Glissando occurs at around 50-th, 90-th, and 170-th segment as a steep change in fundamental frequency and vibrato occurs as a slight fluctuations in fundamental frequency throughout the sound sample. It can be seen from Fig. 3.12 the highest increase in HNR of the ABS-FM is at time intervals where glissando and vibrato (i.e. intervals with the highest frequency modulation) take place.

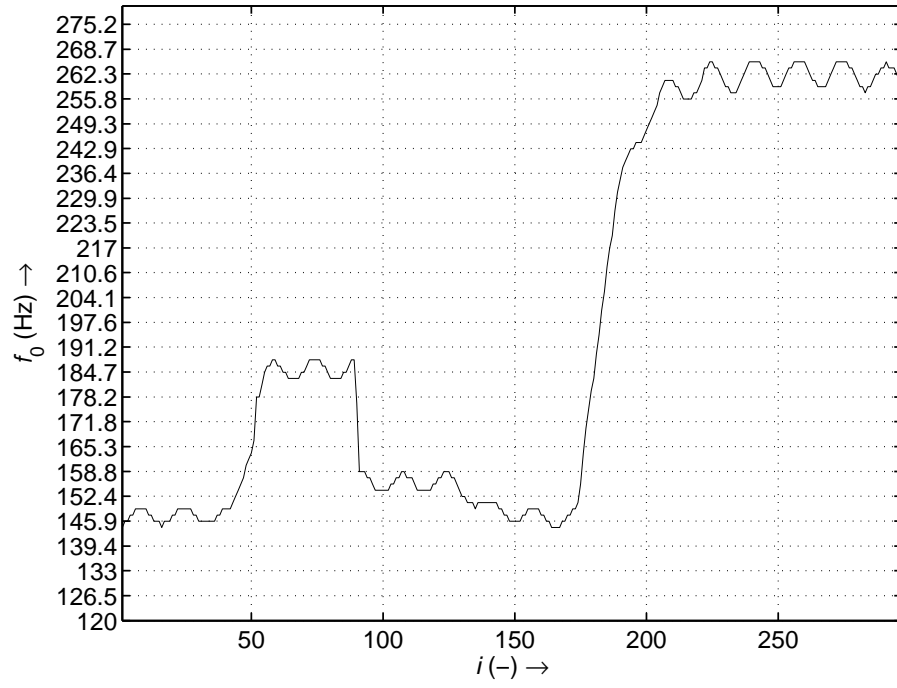


Fig. 3.11: Fundamental frequency of *viola* sound sample.

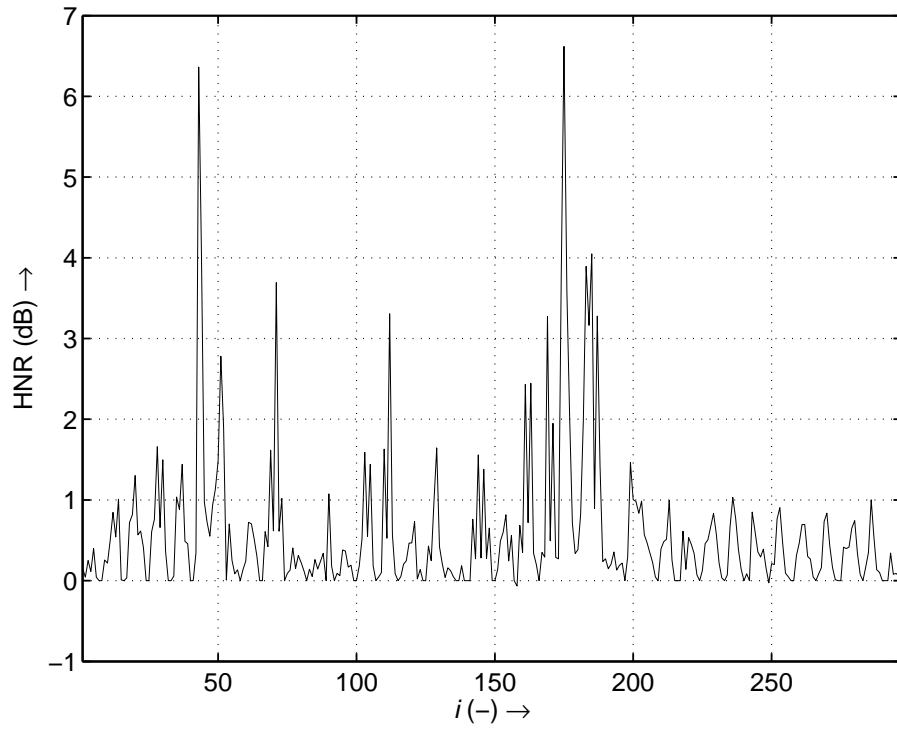


Fig. 3.12: Increase of HNR when using ABS-FM over ABS-S on sound sample *viola*.

4 CONCLUSION

This thesis was focused on methods for representation of harmonic signals with time-varying frequency components. Most of the focus of the methods used is on the Fan-Chirp Transform and Harmonic Transform which are both generalizations of the Fourier Transform for harmonic signals with time-varying frequency components and therefore they share some resemblances.

The chapter 3.1 is dedicated to Harmonic Transform and its computation speed. Fundamental frequency estimation is a prerequisite to computing the Harmonic Transform which has so far been computed using Spectral Flatness Measure. An algorithm to decrease the number of operations needed for SFM computation is presented based on the fact that the Harmonic Transform's image is one-sided. However the Harmonic Transform is enumerated using direct computation from the analytical definition which employs $\mathcal{O}(N^2)$ computational complexity. Therefore further research was aimed at decreasing the computational complexity of Harmonic Transform.

Section 3.2 introduces the Fast Harmonic Transform. The fast transform has been designed by splitting the Harmonic Transform into time-warping of the input signal and performing FFT. This allows for subquadratic computational complexity. However the time-warping operation involves interpolation which introduces noise to the signal and renders SFM ineffective as fundamental frequency change estimation algorithm. It also introduces aliasing which is dealt with using oversampling and different interpolation methods.

Since SFM cannot be used as a fundamental frequency change algorithm for FHT, we have introduced two methods of its estimation. First method is based on computing a gathered log-spectrum on a range of fundamental frequencies and its changes. This method is rather fast though it suffers in fundamental frequency resolution. Second method is based on reconstruction error of harmonic part of the signal using harmonic parameter estimation. This method is slower than the first method, though it offers better resolution in fundamental frequency estimation. Both of these methods have been run on a speech signal *micf01sa02* to compare their results.

Finally, since until now all papers published on the HT have been applied on speech signals sampled at 8 kHz, we wanted to analyze real signals with frequency modulation sampled at 44.1 kHz, which is a common sampling frequency in digital audio. Generally it can be said that the HT decreases reconstruction error (i.e. the ability to represent the signal) for signals with frequency modulation.

BIBLIOGRAPHY

- [1] M. Goodwin and M. Vetterli, “Time-frequency signal models for music analysis, transformation, and synthesis”, in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1996, pp. 133–136.
- [2] G. Wakefield, “Time-pitch representations: acoustic signal processing and auditory representations”, in *Time-Frequency and Time-Scale Analysis, 1998. Proceedings of the IEEE-SP International Symposium on*, Oct. 1998, pp. 577–580.
- [3] A. Makur and S. Mitra, “Warped discrete-fourier transform: theory and applications”, *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, vol. 48, no. 9, pp. 1086–1093, Sep. 2001, ISSN: 1057-7122.
- [4] R. Sluijter and A. Janssen, “A time warper for speech signals”, in *Speech Coding Proceedings, 1999 IEEE Workshop on*, 1999, pp. 150–152. DOI: 10.1109/SCFT.1999.781514.
- [5] R. Baraniuk and D. L. Jones, “Warped wavelet bases: unitary equivalence and signal processing”, in *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP ’93*, 1993, pp. 320–323.
- [6] R. Baraniuk and D. Jones, “Unitary equivalence: a new twist on signal processing”, *Signal Processing, IEEE Transactions on*, vol. 43, no. 10, pp. 2269–2282, 1995, ISSN: 1053-587X. DOI: 10.1109/78.469861.
- [7] L. Cohen, “The scale representation”, *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3275–3292, Dec. 1993, ISSN: 1053-587X.
- [8] T. Irino and R. D. Patterson, “Segregating information about the size and shape of the vocal tract using a time-domain auditory model: the stabilised wavelet-mellin transform”, *Speech Commun.*, vol. 36, pp. 181–203, 3 Mar. 2002, ISSN: 0167-6393.
- [9] A. De Sena and D. Rocchesso, “A study on using the mellin transform for vowel recognition”, in *SMC05, International conference on Sound Music and Computing*, Salerno, Italy, 2005.
- [10] F. Zhang, G. Bi, and Y. Chen, “Harmonic transform”, *Vision, Image and Signal Processing, IEE Proceedings*, vol. 151, no. 4, pp. 257–263, Aug. 2004, ISSN: 1350-245X.
- [13] L. Weruaga and M. Képesi, “The fan-chirp transform for non-stationary harmonic signals”, *Signal Processing*, vol. 87, no. 6, pp. 1504–1522, 2007, ISSN: 0165-1684.

- [14] M. Képesi and L. Weruaga, “Adaptive chirp-based time–frequency analysis of speech signals”, *Speech Communication*, vol. 48, no. 5, pp. 474–492, 2006.
- [15] F. Zhang, Y.-Q. Chen, and G. Bi, “Adaptive harmonic fractional fourier transform”, *Signal Processing Letters, IEEE*, vol. 6, no. 11, pp. 281–283, 1999, ISSN: 1070-9908. DOI: 10.1109/97.796288.
- [16] —, “Adaptive harmonic fractional fourier transform”, in *Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEEE International Symposium on*, vol. 5, 2000, 45–48 vol.5. DOI: 10.1109/ISCAS.2000.857359.
- [17] P. Zubrycki and A. Petrovsky, “Quasi-periodic signal analysis using harmonic transform with application to voiced speech processing”, in *Proceedins of International Symposium on Circuits and Systems (ISCAS 2010)*, Paris, France, May 2010, pp. 2374–2377.
- [18] R. B. Dunn, T. F. Quatieri, and N. Malyska, “Sinewave parameter estimation using the fast fan-chirp transform”, in *WASPAA*, 2009, pp. 349–352.

AUTHOR'S PUBLICATIONS

- [11] M. Trzos, “Frequency warping via warped linear prediction”, in *Telecommunications and Signal Processing (TSP), 2011 34th International Conference on*, Aug. 2011, pp. 348–350.
- [12] —, “Determining the prediction order of warped linear prediction for frequency warping”, *Elektrorevue - Internetový časopis (<http://www.elektrorevue.cz>)*, vol. 2, no. 4, pp. 51–57, Dec. 2011.
- [19] —, “Optimalizace odhadu fázové funkce harmonické transformace”, *Elektrorevue - Internetový časopis (<http://www.elektrorevue.cz>)*, vol. 14, no. 4, pp. 1–4, 2012.
- [20] M. Trzos and H. Khaddour, “Efficient spectral estimation of non-stationary harmonic signals using harmonic transform”, *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, vol. 1, no. 2-3, 2012, ISSN: 1805-5443.
- [21] M. Trzos and H. Ladhe, “Fast implementation of harmonic transform”, in *Telecommunications and Signal Processing (TSP), 2013 36th International Conference on*, Jul. 2013, pp. 498–501. DOI: 10.1109/TSP.2013.6613982.

Curriculum Vitæ

Michal Trzos

- Tel: (+420) 776 690 786
- Email: michal.trzos@gmail.com

Personal information

- Born on 17th June, 1984

Education

- 2009–2015 Ph.D. in Teleinformatics, Brno University of Technology, Brno.
- 2008–2010 M.Sc. in Company Management and Economics , Brno University of Technology, Brno
- 2007–2009 M.Sc. in Communications and Informatics, Brno University of Technology, Brno
- 2004–2007 B.Sc. in Teleinformatics, Brno University of Technology, Brno

Employment

- 2013–present Senior Software Developer, Institute of Computer Science, Masaryk University, Brno, Czech Republic
- 2011–2012 Software developer, Phonexia, Brno, Czech Republic
- 2010–2011 Audio Programmer, Tonic Games, Copenhagen, Denmark
- 2007–2010 Software Developer, Audiffex, Boskovice, Czech Republic
- 2006–2008 ICT Specialist, Masaryk University, Brno, Czech Republic

Participation in Projects

- 2297/G1/2012 – Inovation of the Electroacoustics course. Holder: Ing. Trzos. 2012
- 3160/G1/2011 – Introduction of general-purpose computing on graphics processing units to Multimedia and graphics processors course. Holder: Ing. Trzos. 2011
- FEKT-S-11-17 – Research of Sophisticated Methods for Digital Audio and Image Signal Processing. Holder: Prof. Z. Smékal. 2011
- FEKT-S-10-16 – Research on Electronic Communication Systems. Holder: Prof. K. Vrba. 2010
- FR-TI1/495 – Manifold System for Multimedia Digital Signal Processing. Holder: Ing. Schimmel. 2009–2012

Invited Talks

- Estimation of harmonic parameters of linearly frequency modulated signals using analysis by synthesis approach. 4th SPLab workshop, November 20 to November 21, 2014
- Representation of frequency modulated audio signals using Fast harmonic transform. 3rd SPLab workshop, October 30 to November 1, 2012

Results

- Publications: 13
 - In proceedings of international conferences: 4
 - In other journals: 6
 - In other conferences: 3
- Software/Products: 2
- Citations (without self-citations): 1

Awards

- EEICT 2010 student conference and competition - 2nd prize

ABSTRACT

This thesis deals with representation of non-stationary harmonic signals with time-varying components. Its main focus is aimed at Harmonic Transform and its variant with sub-quadratic computational complexity, the Fast Harmonic Transform. Two algorithms using the Fast Harmonic Transform are presented. The first uses the gathered log-spectrum as fundamental frequency change estimation method, the second uses analysis-by-synthesis approach. Both algorithms are used on a speech segment to compare its output. Further the analysis-by-synthesis algorithm is applied on several real sound signals to measure the increase in the ability to represent real frequency-modulated signals using the Harmonic Transform.

ABSTRAKT

Tato práce se zabývá reprezentací nestacionárních harmonických signálů s časově proměnnými komponentami. Primárně je zaměřena na Harmonickou transformaci a její variantu se subkvadratickou výpočetní složitostí, Rychlou harmonickou transformaci. V této práci jsou prezentovány dva algoritmy využívající Rychlou harmonickou transformaci. První používá jako metodu odhadu změny základního kmitočtu sbírané logaritické spektrum a druhá používá metodu analýzy syntézou. Oba algoritmy jsou použity k analýze řečového segmentu pro porovnání výstupů. Nakonec je algoritmus využívající metody analýzy syntézou použit na reálné zvukové signály, aby bylo možné změřit zlepšení reprezentace kmitočtově modulovaných signálů za použití Harmonické transformace.