A LOW–DELAY ALGORITHM FOR INSTANTANEOUS PITCH ESTIMATION

E. Azarov, M. Vashkevich, D. Likhachov, A. Petrovsky

Computer Engineering Department,
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus

Work was supported by IT ForYou Company and Belarusian Republican Foundation for Fundamental Research (grant No F14MV-014)

1. Introduction

The paper presents an algorithm for instantaneous pitch estimation specifically designed for real-time applications.

The main features of the proposed algorithms are:

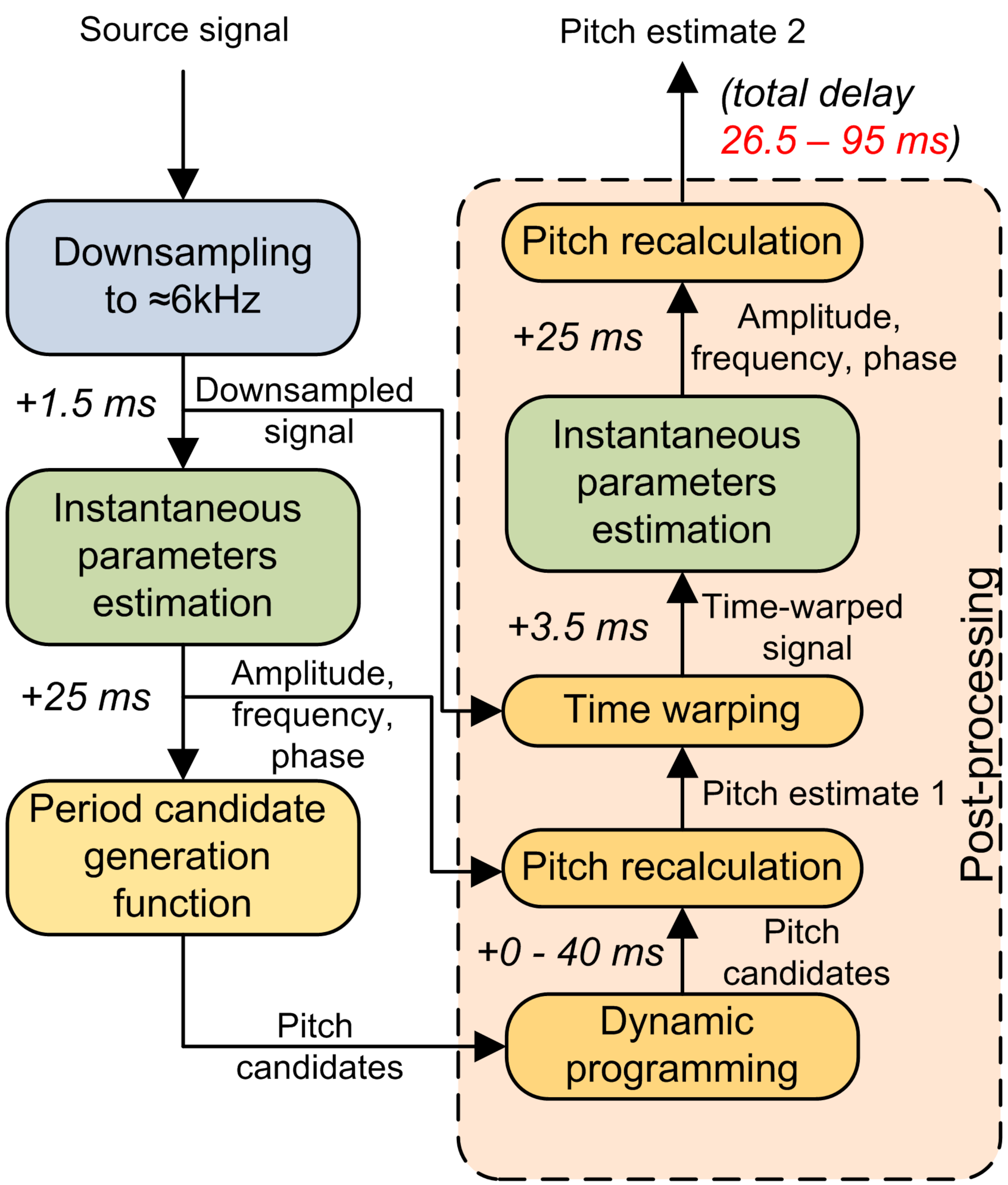
- high estimation accuracy and time resolution;
- low inherent delay;
- efficient processing scheme;
- good performance in noisy conditions;

The general idea:

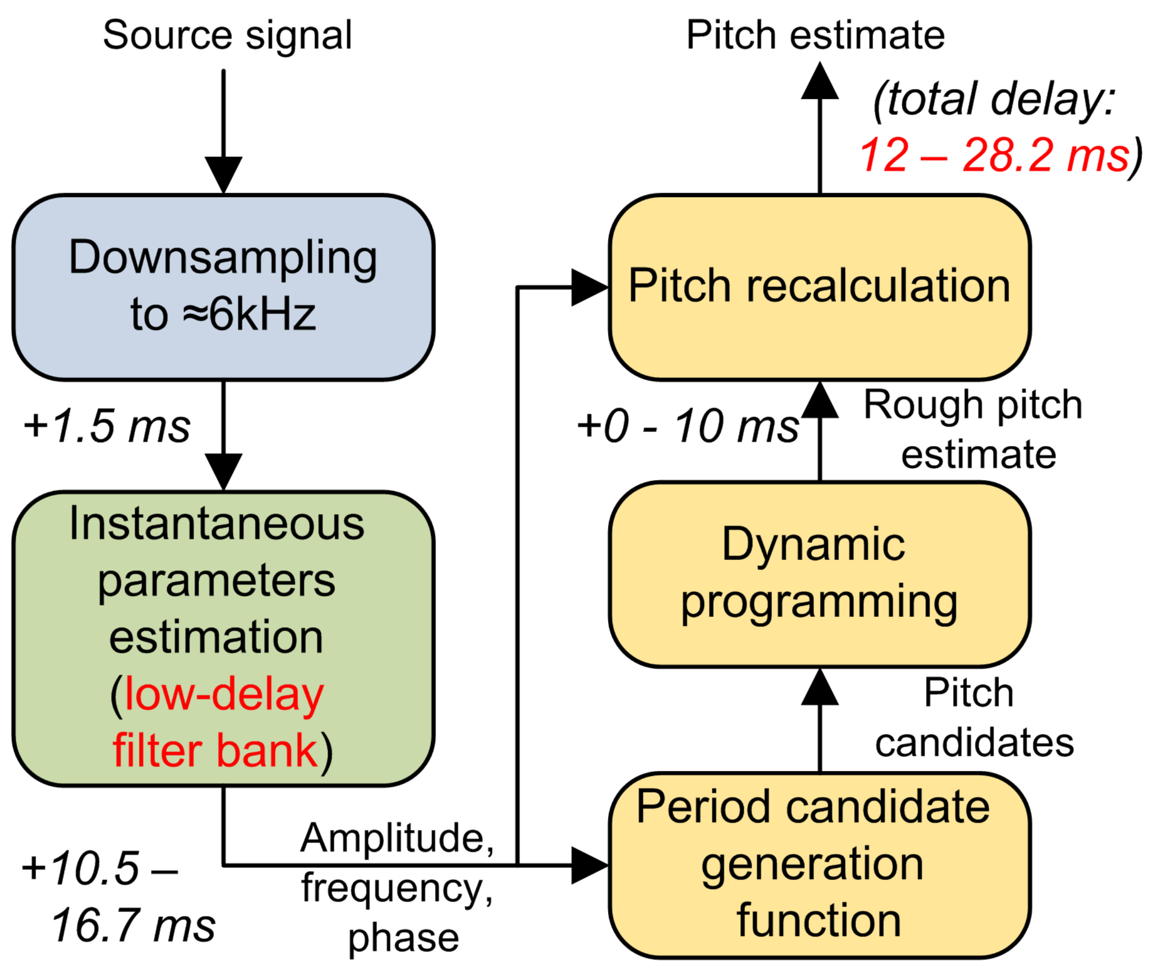
1. the analysis scheme is based on the robust algorithm for instantaneous pitch tracking (IRAPT);
2. a non-linear phase analysis filter bank is applied to get a shorter inherent delay;

2. Pitch estimation schemes

IRAPT scheme



Proposed low-delay scheme



3. Analysis filter bank

The signal $s(m)$ is decomposed into overlapping bandlimited analytical signals $S_{F_\Delta,F_c^i}(m)$:

$$S_{F_\Delta,F_c^i}(m) = \sum_{n=-\infty}^{\infty} \overbrace{\frac{\sin(F_\Delta n)}{n\pi} w(n)}^{\text{linear-phase low-pass filter}} \underbrace{e^{-jF_c^i n}}_{\text{modulation function}} s(m-n) = A_{F_\Delta,F_c^i}(m)\cos\left(\varphi_{F_\Delta,F_c^i}(m)\right),$$

where $2F_\Delta$ - bandwidth and F_c^i - center frequency of the i-th band and $w(n)$ – an even window function. Then instantaneous parameters are evaluated as

instantaneous amplitude $\longrightarrow$ $$A_{F_\Delta,F_c^i}(m) = \sqrt{R^2(m) + I^2(m)},$$

instantaneous phase $\longrightarrow$ $$\varphi_{F_\Delta,F_c^i}(m) = \arctan\left(\frac{-I(m)}{R(m)}\right),$$

instantaneous frequency $\longrightarrow$ $$F_{F_\Delta,F_c^i}(m) = \varphi'_{F_\Delta,F_c^i}(m),$$

where $R(m)$ and $I(m)$ are real and imaginary parts of $S_{F_\Delta,F_c^i}(m)$ respectively.

4. Filter prototype 1 – minimal-phase design

1) Calculate frequency response of the filter: $H_{lp}(k, F_\Delta) = FFT(h_{lp}(n, F_\Delta))$;

2) Calculate cepstrum $C_{lp}(c, F_\Delta) = IFFT\left(\log\left|H_{lp}(k, F_\Delta)\right|\right)$;

3) Set zero for all $c > \frac{N}{2}$ and multiply by 2 except the first:

$$\bar{C}_{lp}(c, F_\Delta) = \begin{cases} C_{lp}(c, F_\Delta), & c = 1 \\ 2C_{lp}(c, F_\Delta), & 1 < c \le \frac{N}{2} \\ 0, & c > \frac{N}{2} \end{cases}$$

4) Calculate minimum-phase freq. response $\bar{H}_{lp}(k, F_\Delta) = e^{FFT(C_{lp}(c, F_\Delta))}$;

5) Calculate impulse response of minimum phase filter form its frequency response $\bar{h}_{lp}(n, F_\Delta) = IFFT\left(\bar{H}_{lp}(k, F_\Delta)\right)$.

5. Filter prototype 2 – IIR approximation

1) Set $\omega_c = 2.64\pi/M$, where $M = 360$, $L = 3$

2) Using low-pass Chebychev filter of type 1 obtain first *N* samples of the impulse response

$$|H(j\omega)|^2 = \frac{1}{1 + \varepsilon^2 T_L^2(j\omega/\omega_c)}$$

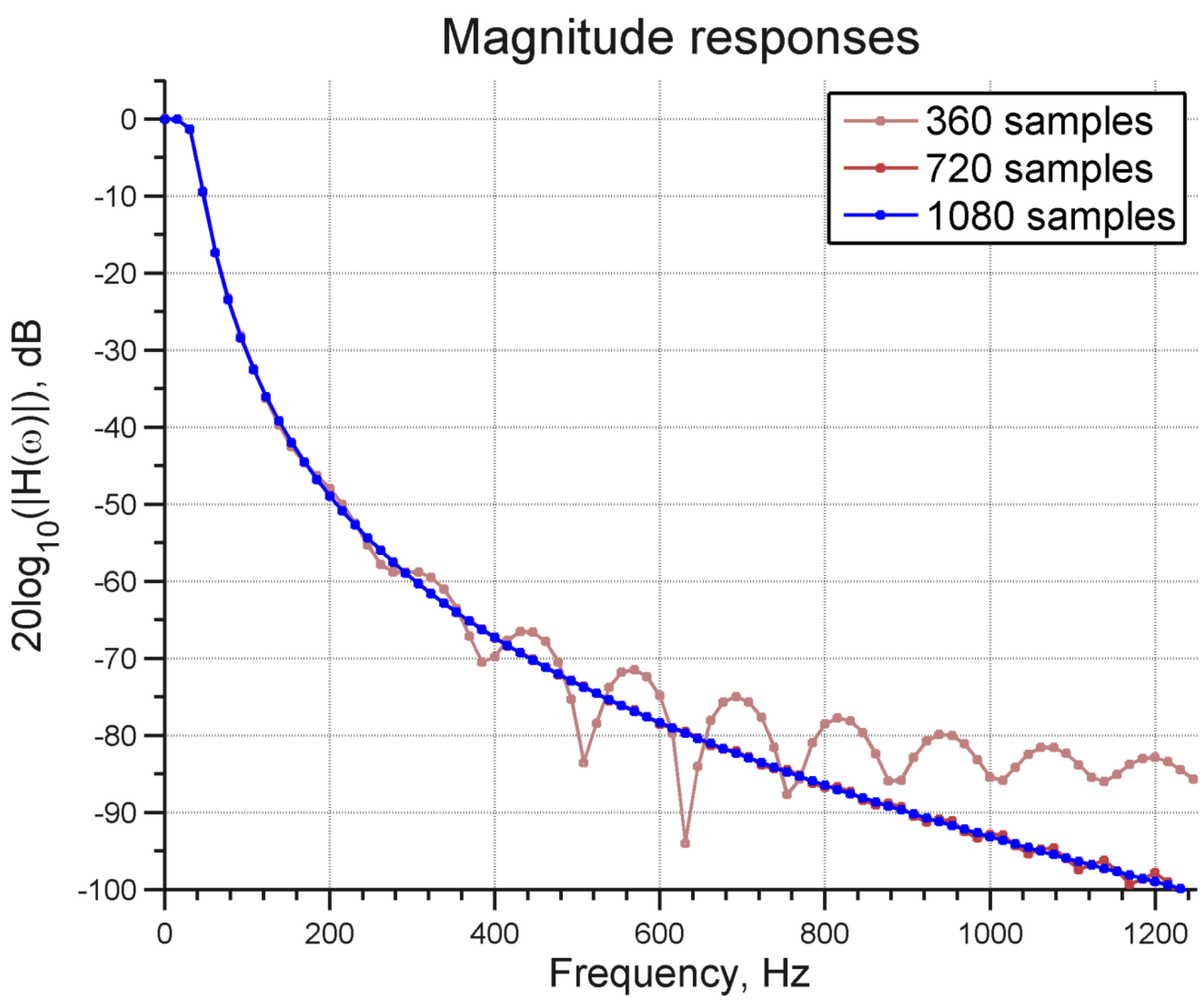
where

$T_L(x)$ – the Chebyshev polynomial of order *L*,

ω_c – the cutoff frequency,

ε – passband ripple parameter.

$N = 1080$ produces a filter with no ripple in the stoppband



6. Comparing filter prototypes

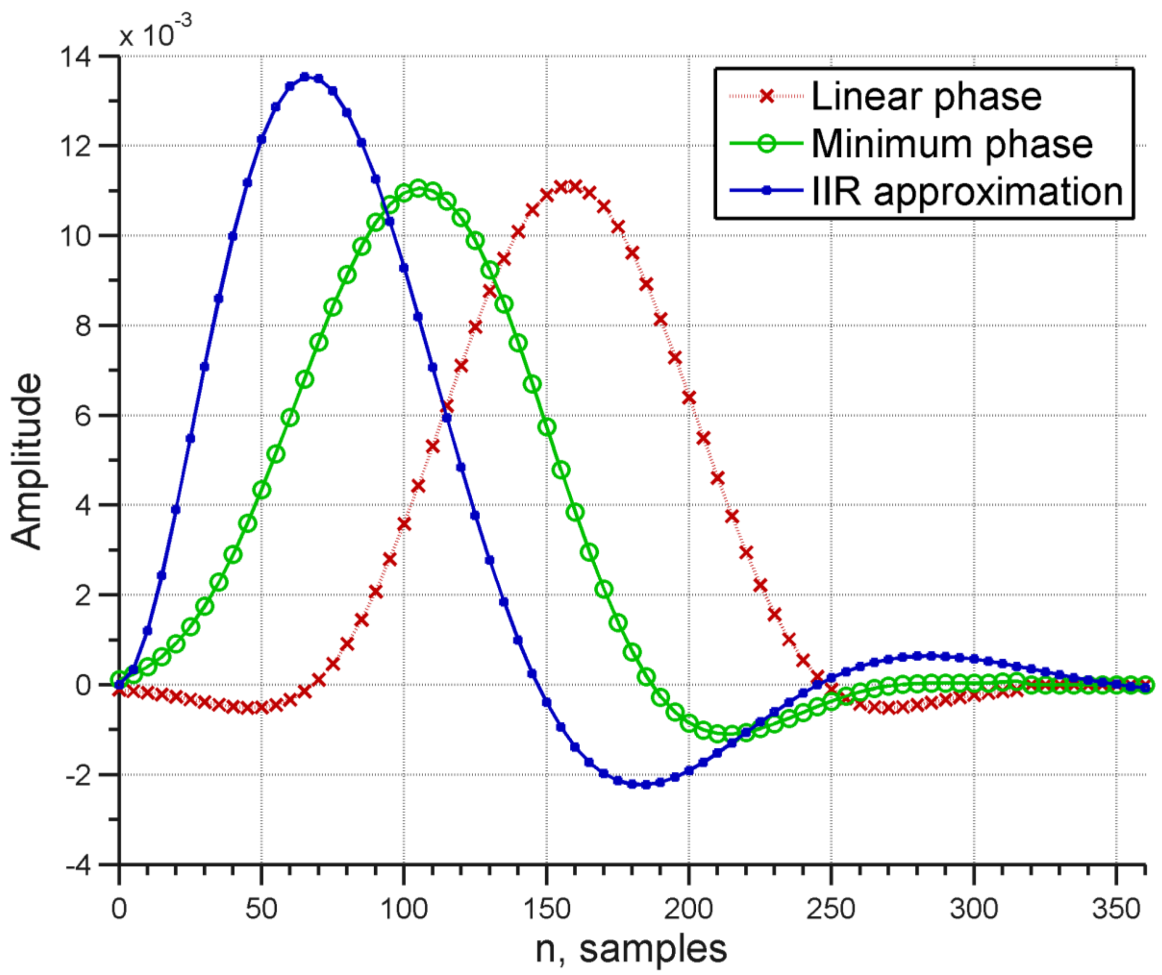
IIR approximation has the shortest response time compared to all competitors.

Linear phase – a constant group delay 25ms

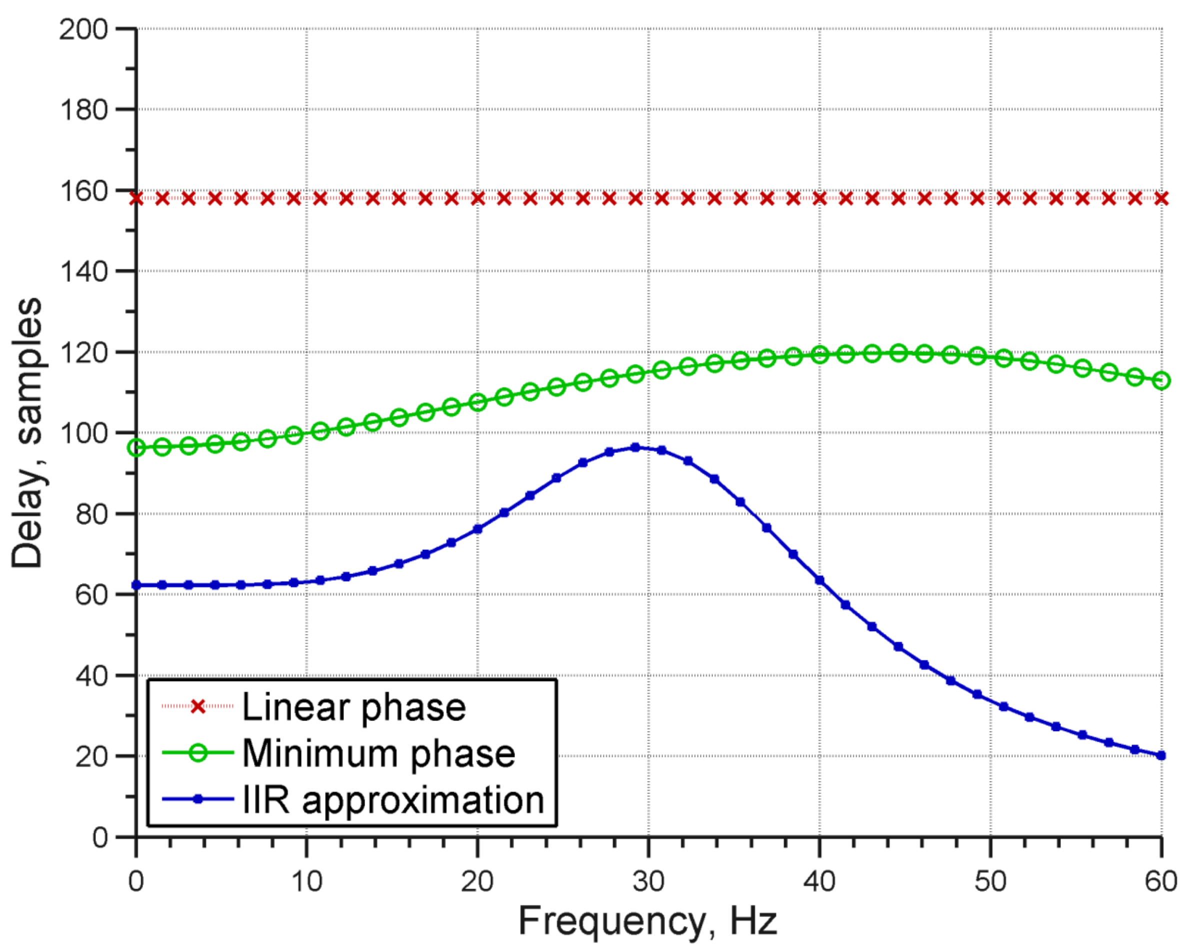
Minimal phase – variable group delay from 16 to 19ms

IIR approximation – variable group delay from 5 to 15 ms

Impulse responses



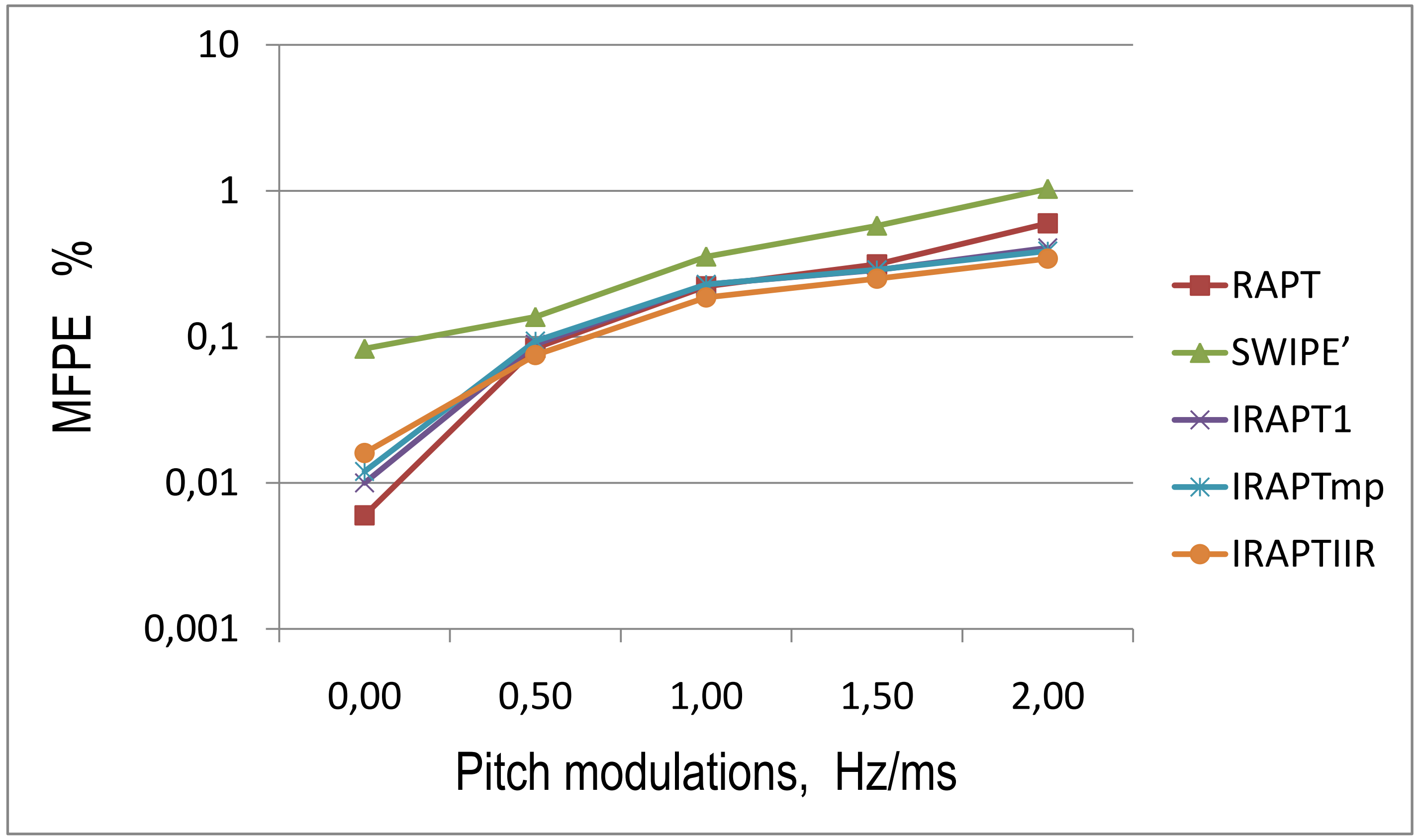
Group delays



7. Experimental results

The proposed technique is compared with other pitch estimation algorithms in terms of gross pitch error (GPE, %) and mean fine pitch error (MFPE, %).

Artificial signals



Natural speech

	Male		Female	
	GPE	MFPE	GPE	MFPE
RAPT	3.69	1.74	6.07	1.18
SWIPE'[4]	0.78	1.51	4.27	0.80
IRAPT 1	1.63	1.61	3.78	0.98
$IRAPT_{mp}$	1.73	1.36	5.82	0.90
$IRAPT_{IIR}$	4.2	1.49	4.83	0.93

[4] A. Camacho and J. G. Harris, "A sawtooth waveform inspired pitch estimator for speech and music", *Journal Acoust. Soc. Am.*, vol. 123, no. 4, pp 1638-1652, Sep. 2008.