A Computationally Efficient Multipitch Analysis Model

T. Tolonen and M. Karjalainen

Gokulraj R (2020102042)   Sushil Kumar Yalla (2020102071)

Aim

* To design a computationally efficient model for multi pitch and periodicity analysis for audio signals.
* We demonstrate a two pitch analysis model by T. Tolonen and M. Karjalainen and how this model is more efficient when compared to the unitary pitch analysis model of Meddis and O'Mard.

Applications / importance of the task

1) In complex audio signals where complex harmonics of various signals are mixed into a single channel it becomes hard to detect it. This can be done by the use of the multi-pitch analysis model.
2) The applications of this analysis include,
   1) Sound source separation
   2) Structural representation of audio signals
   3) Computational auditory scene analysis

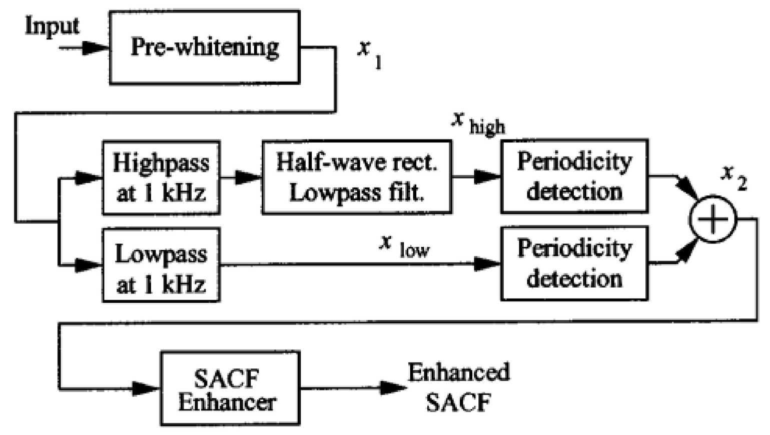
Challenges / Motivation

1) Although the unitary model has shown good correspondence to human perception, it uses multi-channel filterbank which has 32-120 filters. Calculating the auto-correlation for all the channels makes this model computationally inefficient.

2) In this project, we use only 2 channel and also we compute the autocorrelation using DFT. This makes the model computationally more efficient.

Method

* We first use a Hamming window of size 46.4 ms.
* The second step is similar to the unitary model, the signal is passed through two channels one being a high pass filter and another a low pass filter. This high channel signal is then half-wave rectified and lowpass filtered.
* The third step involves periodicity detection which is implemented by using the idea of "generalized autocorrelation". It consists of a discrete Fourier transform (DFT), magnitude compression of the spectral representation, and an inverse transform (IDFT).



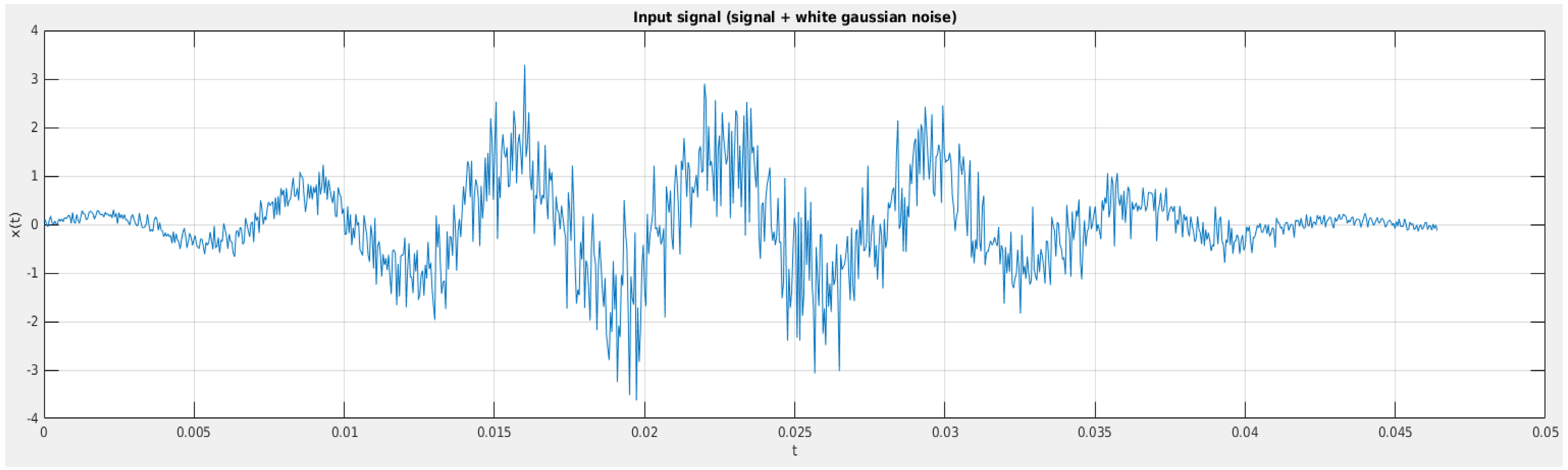
$$x_2 = \mathrm{IDFT}(|\mathrm{DFT}(x_{low})|^k) + \mathrm{IDFT}(|\mathrm{DFT}(x_{high})|^k)$$

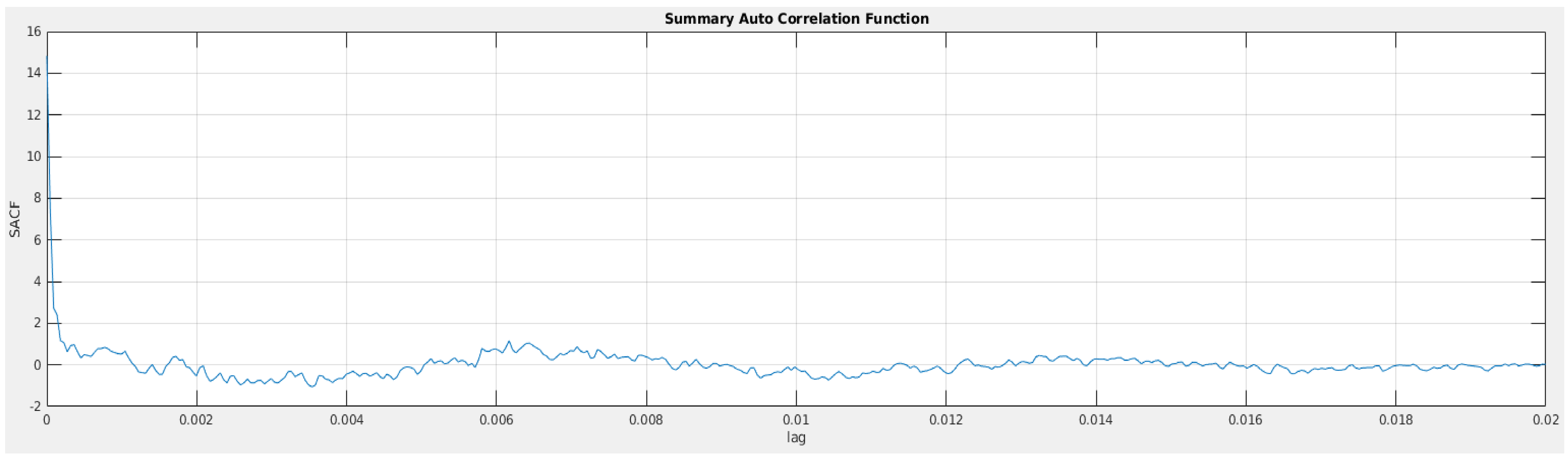
* Finally, we enhance the SACF (i.e., x_2) to obtain ESACF.

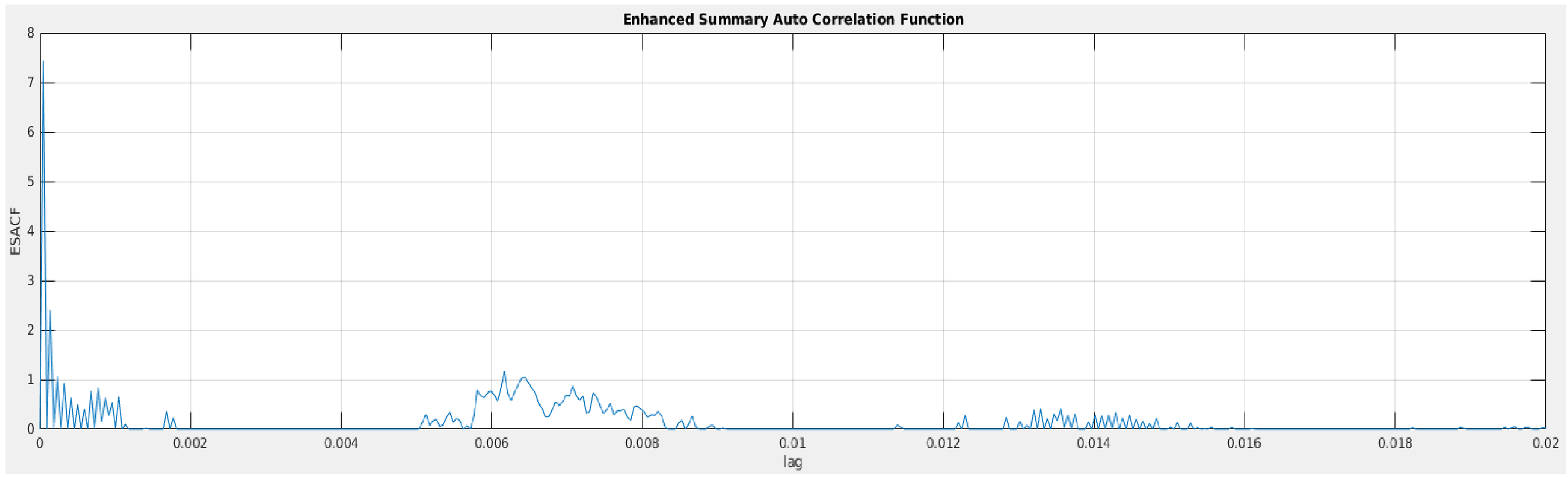
Parameters used in this model

* Window size: 46.4 ms
* Filter order of the high-pass and low-pass: 2
* k = 0.67

MATLAB plots







Conclusion

* Though the auditory analogy of the model might not be strong, this model shows results from the view of human perception, for example, the pre-whitening filter and the channel filter closely associate to certain phenomena in human perception of sound.
* This model shows a much more efficient method of multi pitch analysis which negates the main drawback of the unitary model.

References

[1] Tero Tolonen and Matti Karjalainen. "A computationally efficient multipitch analysis model". In: IEEE transactions on speech and audio processing 8.6 (2000), pp. 708–716.
[2] Ray Meddis and Lowel O'Mard. "A unitary model of pitch perception". In: The Journal of the Acoustical Society of America 102.3 (1997), pp. 1811–1820.