Predictive Analysis of Two Tone Stream Segregation via Extended Kalman Filter

Debmalya Chakrabarty, Mounya Elhilali, *Member, IEEE*

Abstract— **Hearing engages in a seemingly effortless way, complex processes that allow our brains to parse the acoustic environment around us into perceptual sound objects, in a phenomenon called streaming or stream segregation. In this paper, we explore the hypothesis that the auditory system relies on the regularity inherent to each stream to segregate it from other competing streams in the scene. Tracking these regularities is achieved via a recursive prediction that tracks the evolution of each stream, using a Kalman filtering approach. The proposed approach combines spectral analysis operating at the level of the auditory periphery with a temporal analysis using Kalman tracking. To incorporate nonlinear relationships in the signal patterns, we employ an extended Kalman filter. This scheme is tested on sinusoidal patterns, or the two tone paradigm. The combined spectral and temporal analysis developed here is able to predict perceptual results of stream segregation by human listeners in a two tone paradigm.**

Index Terms— **Streaming, Kalman Filter, Sinusoidal Pattern, Alternating, Synchronous, Phase relationships**

I. INTRODUCTION

In everyday acoustic scenes, sounds rarely exist in isolation; and we often experience an assembly of sound events that our brain is able to parse into individual perceptual objects. This ability of the auditory system to parse these intermingled signals into their corresponding perceptual objects has been called "auditory scene analysis" [1]. Understanding how the underlying processes of auditory scene analysis is of great interest to engineers and neuroscientists alike. Emulating the brain's ability to segregate streams can mean promising advances in sound and speech technologies, especially in dealing with nonstationary background distracters and unknown distortions always present in real life environments [2]. It also leads to improved communication aids (hearing aids, cochlear implants, speech-based human-computer interfaces) for the sensory-impaired; as well as a better understanding of the workings of our hearing system.

A number of theories have been proposed about the processes underlying the brain's ability to organize sensory acoustic events into perceptual objects [3]. One of appealing theories has posited that the tonotopic organization of the auditory pathway from the periphery to the auditory cortex underlies that ability of the system to segregate sound elements that activate different frequency channels [4]. The spectral analysis along frequency organized bandpass filters can undoubtedly explain a number of perceptual phenomena related to stream segregation. On the other hand, there has been an argument for the role of temporal patterns and relationships among different sound events in biasing the stream segregation process. Strong evidence suggests that the auditory system relies on the regularity among sound patterns to parse these events into separate or grouped auditory objects [5]. An obvious example in this case is onset synchrony which is believed to play an important role in grouping sound elements together over both short and long time scales [6].

In the current work, we attempt to explore the role of both this spectral analysis and temporal tracking in a simple of two-tone paradigm. This paradigm is one of the simplest and most compelling demonstrations of auditory streaming, which involves tone sequences such as those shown in Figure 1 The tones are either heard as a single stream with a distinct *galloping* rhythm (upper panel), or as two separate perceptual streams of regular tones (lower panel). Whether the sequence is perceived as one or two streams depends on a number of variables, including the frequency and temporal separation of the A and B tones, the overall duration of the sequence, and the attentional set of the listener [7] [8] [9] [10]. As the two tones become further apart (bigger ΔF), they are easier to segregate in a clear demonstration of the role of peripheral channeling in activating different frequency channels, hence

D. Chakrabarty and M. Elhilali are with the department of Electrical and Computer Engineering, Laboratory of Computational Audio Perception, Center for Speech and Language Processing, Johns Hopkins University, MD 21218, USA. Corresponding author: M. Elhilali, phone: 410-516-8185, email: mounya@jhu.edu.

978-1-4673-5239-0/13/$31.00 ©2013 IEEE

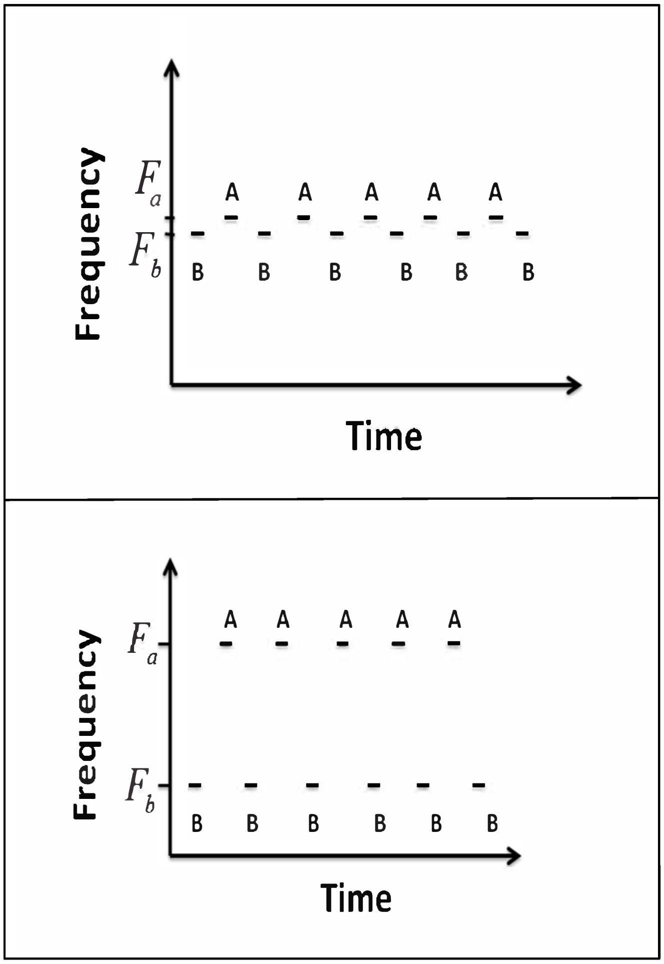

Figure 1: Upper Panel: Single stream with *galloping* rhythm. Lower panel: Two streams with regular tones A and B.

facilitating the ability to separate the two tone sequences into separate streams [11]. However, if one changes the timing relationship between these two tone sequences and makes them in synchrony, the system uses this synchrony as indication that they are to be grouped together, despite the large ΔF separation between them [12]. This later is evidence of the role of the temporal analysis in stream segregation.

In the current paper, we explore both types of analyses using this two tone paradigm. The spectral analysis is achieved via a filterbank decomposition mimicking the analysis at the level of the auditory periphery [17]. The temporal analysis is achieved via temporal tracking of the regular sound patterns. The later involves following the evolution of the temporal patterns in an online fashion. If regular over time, these patterns could be predicted with an appropriate model and hence tracked with high degree of fidelity. We implement this tracking process using a predictive Kalman filter approach. The choice of this method is motivated by both its non-deterministic nature; as well as its recursive, online formulation that integrates the prediction and estimation stage symbiotically. The current implementation uses an extended Kalman filter to allow for tracking sinusoidal patters such as pure tones.

The paper is organized as follows. In section II, we are going to discuss about the whole system in detail. We are going to look into each of the blocks which comprise the system. In Section III, we will be discussing the results which will be followed by the conclusion in Section IV.

II. System Architecture

In this section, we discuss in detail about the Kalman filter-based predictive system and each of its components as shown in Figure 2.

A. **Signal**: The signal used in this context is a two tone sinusoidal signal. The sinusoidal signal is represented as:

$$x(t) = Sin(2\pi(f_a + \phi)t) \qquad (1)$$

where f_a is the carrier frequency and ϕ is the phase associated with the sinusoidal tone respectively. The two tone paradigm involves manipulating the frequency as well as phase relationship between two tone sequences. We have used a silence region of 10 ms between the two tones. In the current paper, we fix the duration of silence region between the tones and manipulate the carrier frequencies as well as the phase.

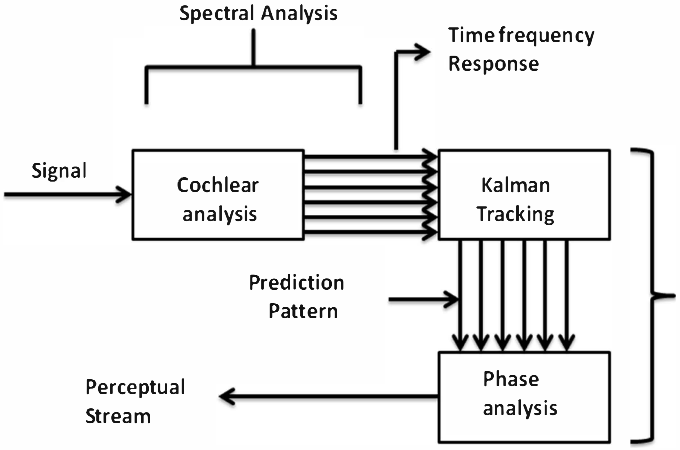

Figure 2 : System Architecture showing each of the components that make the Kalman filter based predictive system.

Based on the choice of carrier frequency and phase relationship between the two tones, we explore 4 types of signal:

1. Alternating pattern with large difference between the carrier frequencies.
2. Alternating pattern with small difference between the carrier frequencies.
3. Synchronous pattern with difference between the

carrier frequencies.

4. Synchronous pattern with difference between the carrier frequencies.

Our idea is to test how the Kalman filter tracks the regular temporal patterns associated with these alternating and synchronous patterns and how can we correlate it with the way brain tracks these patterns and do the streaming.

B. **Cochlear Analysis**: In cochlear filterbank analysis, we have made use of a bank of gammatone filters defined by Patterson and Holdworth [13] for simulating cochlea filtering. The Gammatone Filters are widely used in the modeling of auditory system and can effectively mimic the the analysis at the level of the auditory periphery. The impulse response of Gammatone filter [14] is given as

$$g(t) = at^{n-1}\cos(2\pi ft + \phi)e^{-2\pi bt} \qquad (2)$$

Where **n** is the order of the filter, **b** is the bandwidth of the filter, ϕ is the phase, **a** is the amplitude and **f** is the filter center frequency.

In this case, we have used a set of 50 overlapping Gammatone Filters through which the input signal of two tone stream is passed. The time frequency response of the filterbank is then passed through a low pass filter, The main purpose for using this low pass filter is to retain only the envelope characteristics of the input and throwing away all the fast varying frequency responses. The reason for doing is that the temporal pattern is a slow varying frequency response and contains only the envelope information [18]. The low pass filter used in this case is 4th order Butterworth Filter with a cutoff of 100 Hz.

This cochlear filterbank analysis gives the time frequency analysis of the signal or the auditory specgtrogram. The time frequency response is then passed to the Kalman filter block.

C. **Kalman Tracking**: The Kalman filter is a very efficient way of predicting and estimating a non deterministic signal [19]. Basic Kalman Filter is a linear estimator which tries to estimate the state of the signal from all past observations. The signal to be estimated and the observation sequence are expressed as a linear state space equation. Then there is a predication and an estimation stage which actually estimates the present state of the signal. The extended Kalman Filter (ECKF) [20] implemented in this study is used to predict and track the regular temporal patterns of the time frequency response of the cochlear filterbank. Because of the use of sinusoidal patterns, the pattern cannot be formulated in a linear form, hence the need for the extended Kalman Filter. The basic equations used in the extended Kalman Filtering approach are given below [15]:

$$y_k = Hz_k + v_k \qquad (3)$$

where $z_k = \sin(2\pi f_i t_k)$, k=0,1,2,3.......N and f_i is the frequency of the i^{th} sinusoid and H=[0 1]. Equation (3) is the state space equation which relates the observation sequence y with the signal to be estimated along with white gaussian noise.

The main idea behind Kalman based tracking is that Kalman filter estimates the present state by making use of all the past states of the observation signal y; which is the predictive nature of this approach.

$$z_{k+1} = f(z_k) \qquad (4)$$

The crux of the filtering lies in the above equation where the future output is represented as a function of the past output hence, it becomes a function of the observation signal y_k as well upon which the rest of the equations depend.

$$\hat{z}_{k|k} = \hat{z}_{k|k-1} + K_k(y_k - H\hat{z}_{k|k-1}) \qquad (5)$$

$$\hat{z}_{k|k+1} = f(\hat{z}_{k|k}) \qquad (6)$$

The above equation is actually the estimation equation which is trying to estimate the present state of the signal z_k from the past states of the signal z_{k-1} and the observation signal y_k. K_k is the Kalman Gain which is a function of the past values of prediction error $\hat{P}_{k-1}$ is shown in equations (7).

$$K_k = \hat{P}_{k|k-1}H^{*T}[H\hat{P}_{k|k-1}H^{*T} + 1]^{-1} \qquad (7)$$

As we can see from equation (7), K_k is estimated from the past states of the prediction error $\hat{P}_{k-1}$. This K_k is then used in estimating the present state of the predication error $\hat{P}_k$ as shown in equation (8). Hence the idea is that at each iteration, the previous value of prediction error is involved in finding the present error, since the filter is adaptive, and the error keeps on decreasing at each iteration. When the optimum number of iterations is reached, the prediction error almost goes to zero.

$$\hat{P}_{k|k} = \hat{P}_{k|k-1} - K_k H\hat{P}_{k|k-1} \qquad (8)$$

$$\hat{P}_{k+1|k} = F_k\hat{P}_{k|k}F_k^{*T} \qquad (9)$$

$$F_k = \partial f(z_k) / \partial z_k \qquad (10)$$

In the above equations, k is the present state, k-1 is the previous state and k+1 is the future state which validates the fact that the Extended Kalman Filter approach is a predictive and iterative approach which keeps on estimating the present state by predicting from past observations. K is the Kalman Gain, $\hat{P}$ is the error covariance matrix, H is the standard observation matrix and F is the taylor series factor that transforms a non-linear function to the linear function over which all the equations of a basic Kalman Filter can be applied. Hence F is the function which is essential for the concept of Extended Kalman Filter. Another important thing that matters while implementing Kalman Filter is the proper selection of initial state for the estimator $\hat{z}_{-1}$ and the prediction error $\hat{P}_{-1}$. The performance of ECKF depends a lot over these initial conditions [15].

This Kalman block produces the prediction tracking patterns from the time frequency response. Finally, the last stage is to find the phase relationship between these tracked patterns to find out whether the two tone (synchronous or alternating) paradigm belongs to single source or two separate sources.

D. **Phase relationship**: The tracked patterns are sent to the phase extraction block which computes the phase difference ϕ between the tracked patterns and compares it with a threshold ζ and gives the decision based on following criteria:

If $\phi < \zeta$, then two tones belong to single source, else two sources are present, where ζ is set equal to pi/10 radians.

III. Results

In this section, we discuss the system's behavior using the two tone paradigm. The spectrogram analysis of time frequency response for alternating and synchronous tones obtained after the cochlear filterbank analysis stage is shown in Figure 3, 4, 5 and 6. In Fig 3 and 4, we can easily discern that two different channels are active in a different manner for alternating and synchronous tones when the carrier frequency difference between the two tones is very high. Now the question arises how the Kalman Filter is going to track this pattern generated by the alternating and synchronous tones? How big is the phase difference between the two tracked patterns for alternating and synchronous case? Fig 5 and 6 shows the spectrogram analysis for two alternating and synchronous tones with a very small difference between the carrier frequencies. On looking into these spectrograms, we can easily say that the two streams belong to a single source and hence form a single stream. Hence we don't need to go the stage of tracking as it is clear enough from the spectral analysis that when the frequency difference between the carriers is very small, the tone pattern forms a single stream.

The next set of results is the tracking patterns which we got from the Kalman tracking block in Figure 2. These results are shown in Figures 7 and 8 respectively. From Figures 7 and 8, we can see that the Kalman Filters take some time to settle down i.e., it takes time for the buildup before it actually starts tracking the actual phase which validates the fact the streaming takes time to buildup hence we can correlate between the tracking pattern of Kalman Filter and the actual streaming which the brain performs.

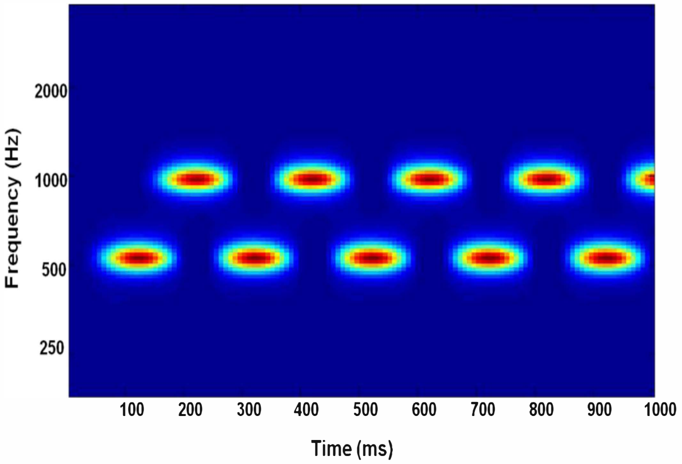

Figure 3: Spectrogram analysis of alternating tones with a large frequency separation

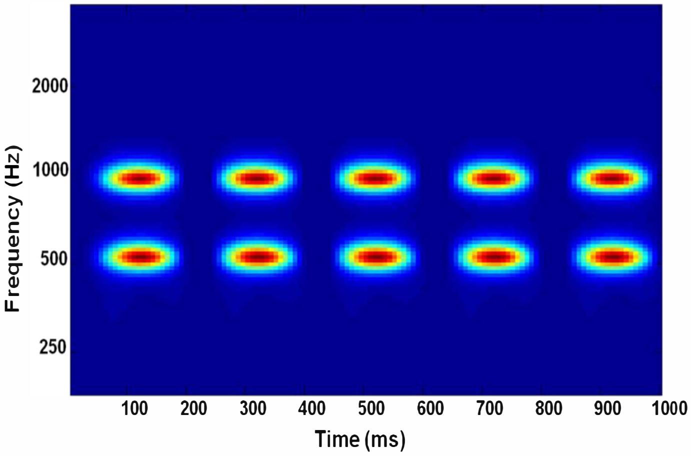

Figure 4: Spectrogram analysis of synchronous tones with a large frequency difference

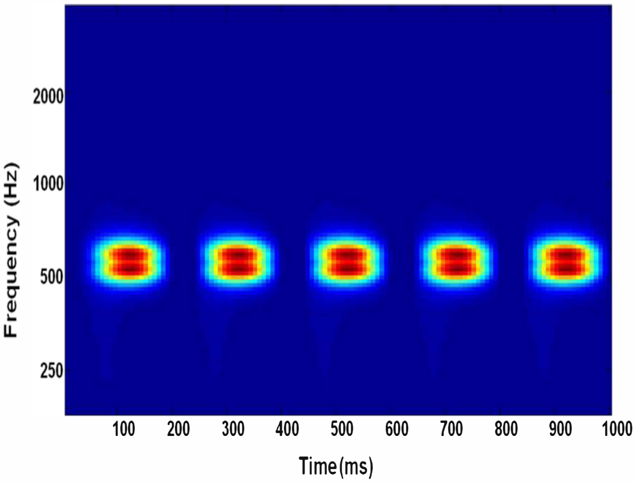

Figure 5: Spectrogram analysis of synchronous tones With a small frequency difference

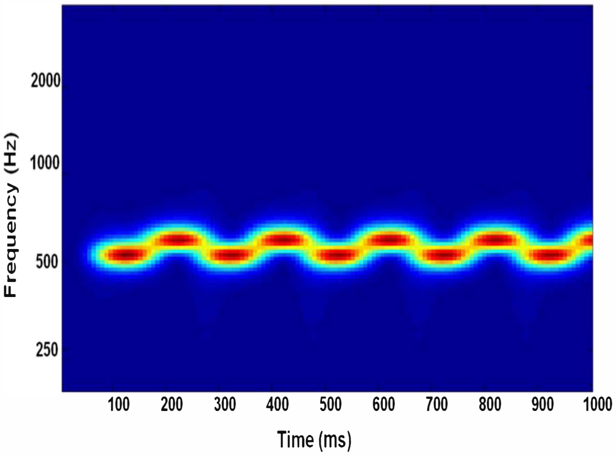

Figure 6: Spectrogram analysis of alternating tones With a small frequency difference

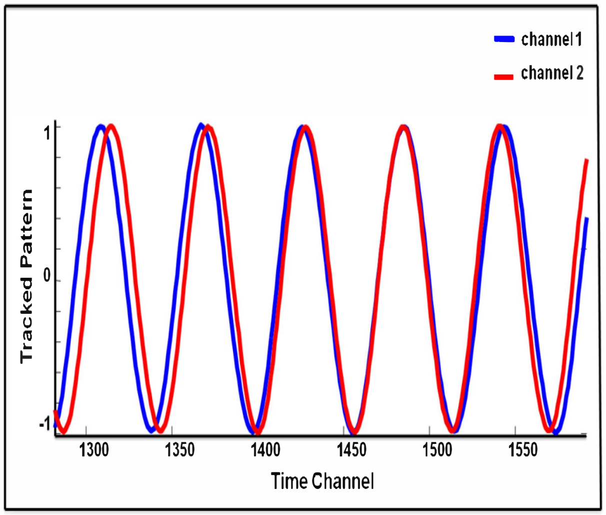

Figure 7: Tracking Pattern for synchronous tones with large difference between carrier frequencies.

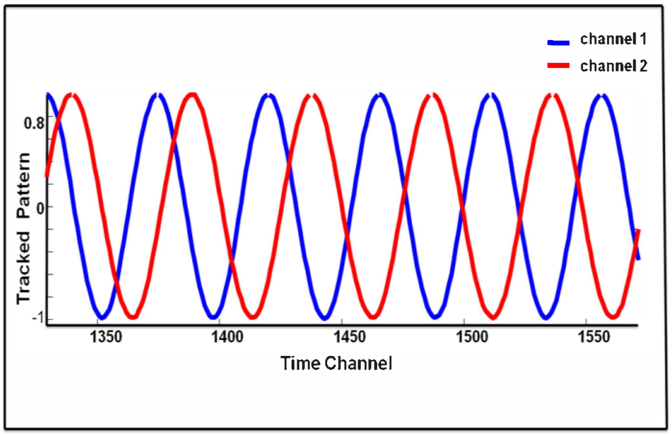

Figure 8: Tracking Pattern for alternating tones with large difference between carrier frequencies.

When these two tracking patterns are sent to the phase extraction block, we calculate the phase difference between the two in both the cases (alternating and synchronous). The phase difference in both the cases is reported in Table 1.

Tone pattern	Phase Difference (ϕ)
Alternating Tone	0.57 radians
Synchronous Tone	0.19 radians

Table 1: Phase difference between the tracked patterns for alternating and synchronous tones.

From Table 1, we can see that the phase difference between the tracked patterns in case of alternating tones is very high than the threshold value (ζ =pi/10 radians), hence the alternating tones with high difference between the carriers easily get segregated into two streams whereas synchronous tones always get streamed as a single stream because of very less difference between the phase of tracked patterns. This observation validates the fact that when there is a large frequency difference between two alternating tones, they always get streamed into two streams whereas synchronous tones always form a single stream whatever the frequency difference between the tones may be because of the in phase criterion [12]. Hence these results suggest that the predictive Kalman Filter tracking of temporal pattern from the time frequency response of the cochlear filter banks correlate with the way our brain tracks the temporal patterns and perform the streaming [16].

IV. Conclusion

The current study reinforces the role of both spectral and temporal processing in stream segregation; it also strengthens the claim that prediction is very much a part of perception. Perceptual objects emerge as information from past sensory cues is integrated over time and able to predict the evolution of its regularity. The Kalman filter is well suited to implement such online tracking mechanisms, since it assumes an underlying Markov model. This work is tested on a very simple paradigm; but lays the foundation to exploring more complex acoustic inputs as well as exploring integrating other processes known to play a role in the process of auditory scene analysis.

ACKNOWLEDGEMENT

This research was supported by NSF grant IIS-0846112, NIH grant 1R01AG036424 and ONR grants N000141010278 and N00014-12-1-0740.

References

[1] Bregman, A.S. (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound*, Bradford Books, MIT Press.

[2] S Greenberg, A Popper, W Ainsworth*, Speech Processing in the Auditory System*, Springer, Berlin, 2004

[3] Istva´n Winkler, Susan L. Denham, Israel Nelken,(2009) "*Modeling the auditory scene:predictive regularity representations and perceptual objects*," Trends in Cognitive Science., Vol.13, Issue 12, 532-540.

[4] Fishman, Y.I., Reser, D.H., Arezzo, J.C., and Steinschneider, M. (2001). *Neuralcorrelates of auditory stream segregation in primary auditory cortex of the awake monkey*. Hear. Res. 151, 167–187.

[5] Winkler, I. et al. (2005) *Event-related brain potentials reveal multiple stages in the perceptual organization of sound*. Brain Res. Cogn. Brain Res. 25, 291–299.

[6] Liang, H., Bressler, S.L., Ding, M., Desimone, R., and Fries, P. (2003). *Temporal dynamics of attention-modulated neuronal synchronization in macaque* V4.Neurocomputing 52-54, 481–487.

[7] Bregman AS, Campbell J (1971) *Primary auditory stream segregation and perception of order in rapid sequences of tones*. J Exp Psychol 89:244-249.

[8] Miller GA, Heise GA (1950) *The trill threshold*. J Acoust Soc Am 22:637-638.

[9] Bregman AS, Ahad P, Crum PA, O'Reilly J (2000) *Effects of time intervals and tone durations on auditory stream segregation*. Percept Psychophys 62:626-636.

[10] Van Noorden, L.P.A.S. (1977) *Minimal Differences of Level and frequency for perceptual fission of tone sequences ABAB*. J.Acoust.Soc.Am. Vol. 61, 1041-1045.

[11] Brain C.J Moore, Hedwig Gockel (2002) *Factors influencing Sequential Stream Segregation*. Acta Acustica United With Acustica, Vol. 88. 320-332.

[12] Shihab A. Shamma, Mounya Elhilali, Christoph Micheyl (2011) "*Temporal Coherence and attention in auditory scene analysis*". Trends in Neurosciences,Vol. 34, No.3, 114-123.

[13] Slaney, Malcolm (1993). "*An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank*". Apple Computer Technical Report #35.

[14] Hohmann, V., *Frequency analysis and synthesis using a Gammatone filterbank*, Acta Acustica united with_Acustica, Volume 88, Number 3, May/June 2002 , pp. 433-442(10**).**

[15] Kiyoshi Nishiyama, *A Nonlinear Filter for Estimating a Sinusoida Signal and Its Parameters in White Noise: On the Case of a Single Sinusoid*, IEEE Transactions on signal processing, Vol. 45, No. 4, April 1997.

[16] Patel D. Aniruddh, Balaban Evan, *Temporal Patterns of human cortical activity reflect tone sequence structure*, *Nature* 404, 80-84 (2 March 2000)

[17] Patterson, R.D. and Moore, B.C.J. (1986) *Auditory filters and excitation patterns as representations of frequency resolution*. In: Moore, B.C.J. (Eds), Frequency Selectivity in Hearing. Academic Press Ltd., London, pp. 123-177.

[18] Chi, T., Gao, Y., Guyton, M.C., Ru, P., and Shamma, S. (1999). *Spectrotemporal modulation transfer functions and speech intelligibility*. J. Acoust. Soc. Am. 106, 2719–2732.

[19] R. R. Bitmead, A. C. Tsoi and P. J. Parker, "*A kalman filtering approach to short-time Fourier analysis*," *IEEE* Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 1493–1501, June 1986.

[20] P. J. Parker and B. D. O. Anderson, "*Frequency tracking of nonsinusoidal periodic signals in noise*," *Signal Processing*, vol. 20, pp.127–152, 1990.