International Journal of Scientific Research in Computer Science and Engineering
Review Paper Volume-3, Issue-3 ISSN: 2320-7639

Review on Multi Pitch Detection in Speech

Rupinder Kaur[1] and Navdeep Kumar[2*]

[1,2*]Dept. of Computer Science and Engineering
Chandigarh University - India

Available online at www.isroset.org

Received: 04 May 2014 Revised: 16 May 2014 Accepted: 08 Jun 2015 Published: 30 Jun 2015

Abstract- This paper includes the review of research being carried on multiple pitches detection in an audio signal. It includes determining multiple fundamental frequency or different pitched sounds. It discusses different algorithms used to identify different speeches spoken at one time. Although, this technique is used for various applications of speech recognition, but it is widely used in music transcription.

Keywords- Speech Recognition, MFCC, LPC, Pitch Tracking, Multiple Pitch Estimation

I. INTRODUCTION

Speech is an analog signal which is further processed in order to attain the desired results. The analog signal is first converted to digital signal through quantization and sampling. This digital signal then undergoes some transforms and thereby fundamental frequency and pitch is detected. However, speech undergoes different phases for different processes. This includes feature extraction, recognition model, distance measures etc. but for multi pitch detection, we shall use feature extraction techniques along with some other algorithms. These algorithms are used to process different pitches in same audio signal.

This paper includes speaker dependent working of speech signals. Speech research is based on various important factors including accuracy, varying bit rates, varying pitch, size of data set. Speech corpus is database of speech signals. It contains various speech recordings.

II. MULTI PITCH DETECTION

There are multiple algorithms used to detect multiple pitches in speech. These include the study of fundamental frequency and spectral features. Multiple pitch detection is important real time scenario. This is used by human being to identify what multiple people are speaking at one time. Related to multi speech, it is important to know how many pitches are there in one signal. Earlier, this work was performed for multiple pitches of music and instrumentations. But now, it has been extended to human speech detection. If an indifferent pitch is detected, one can identify spectral component and vice versa. These spectral components are believed to be dependent on various acoustic cues.

III. WORKFLOW OF ALGORITHMS

In 1964, A. M. Noll [1] introduced the cepstral analyzer. This has been through short time power spectra and band-pass filter bank process. This cepstral is used to determine the fundamental frequency and determine the voiced-unvoiced signals in audio wave. It has been designed on IBM digital computer. However, he in 1966 [2] introduced the advanced technique of computing power spectrum of log of power spectrum to obtain the better peak. Th estimation of pitch was much better than the previous experiment. This experiment has been performed at Bells Laboratory and cepstrum of the signal was calculated and automatically gets plotted on microfilm. In 1976, L R Rabiner [3] performed analysis of eight pitch detection algorithms. This experiment has been performed on considerable size of speech corpus. The errors are also measured and relative deviation is being observed for the same.

In 1999, P. J. Walmsley et al [4] proposed Bayesian probabilistic framework for pitch detection. This estimates harmonic model parameters as it work upon correlation between adjacent frames and variation of frequency over time. M Karjalainen et al [5] introduces the concept of estimation of pitches in audio signal and thereby separating different source signals from complex speech signals. T. Tolonen et al [6] presented the model for computing multiple pitches from speech signal by diving it into two channels namely low channel and high channel. This is followed by auto correlation method. This is used for real time scenario.

In 2001, A. P Klapuri [7] proposed a new model after implementation of three models for multi pitch detection in

speech signals. This work contains spectral smoothness evaluation. The speech corpus being used varies from one to six speeches in an audio wave. The error rate reduces as we reduce the number of sounds in one audio wave. In 2003, M Wu et al [8] used HMM for detection of pitch tracks being framed. The robust algorithm is proposed for speech recognition algorithm for noisy speech. A. P Klapuri [9] in his next paper, performed another experiment in which he calculates harmonicity and spectral smoothness. This is repetitive process in which one sound is being detected first. This sound is then removed from complex mixture of sounds in signal. Now, the residual signal is undergone through same process.

In 2003, S. S. Abeysekera et al [10] proposed a new technique for the same using Bispectrum that means two dimensional frequency log. This undergoes detection of one signal and removing it from the testing signal. Same process is being performed for residual signal. Removal of two dimensional signal is comparatively easy and convenient. M. G. Christensen et al [11] introduced some solutions for separation of speech signals from MUSIC.

In 2008, X Zhang [12] introduced an algorithm based on weighted summary correlogram. Accuracy measures are observed better. A. Klapuri [13] performed better accuracy with fundamental frequency detection and removal of sound one by one iteratively. R. Badeau et al [14] proposed new expectation maximization algorithm for the same. It also separate the overlapped harmonic spectra obtained from speech signals.

In 2010, E. Vincent et al [15]proposed models for time varying amplitude speech signals. They used their basic model as non-negative matrix factorization. However, E Benetos et al [16] also worked on time-varying multi pitch detection in speech signals. They used HMM for speech corpus containing MIDI database. A. Koretz et al [17] used Maximum A posteriori probability algorithm for multiple pitch detection. S. I. Adalbjornsson et al [19] worked on block sparsity. They introduced an alternative algorithm for the detection of multiple pitches in speech signal. The work is going on the the similar concept and new techniques have been introduced. This is further carried on to better accuracy and size of dataset.

Wohlmayr *et al.* [20] proposed a methodology to estimate the stream pitches of simultaneous talkers by using factorial hidden Markov approach. Initial training of system is done by using the isolated recordings mixtures of talkers. This is the basic limitation of the systems where initial training on sources is not feasible. [20], [22].

Recently an unsupervised approach for estimating the stream pitches is proposed by Hu and Wang [30]. The approach focuses on separating the signals of two simultaneous talkers. The approach is tested only for speech and its applicability for other audio signals such as music is not tested.

In 2011, Mads Graesboll et al. [23] the task of high resolution pitch estimation is motivation of the research of this paper. Already proposed systems such as the classical comb filtering, maximum likelihood methods and others based on optimal filtering are extended for unknown number of harmonics. This is also known as model order which is based on the posteriori principle. So this proposed method of estimating orders and fundamental frequency is applicable to those situations also where there is no prior knowledge of the model order. Also a computationally efficient order-recursive implementation that is much faster than a direct implementation has been proposed.

In 2012, John Xi Zhang et al. [24] extends the multi-pitch estimation to a level of multi-channel. A new estimator is proposed which deals with estimating fundamental frequency as well as DOA of multiple sources. Subspace analysis along with time space model is used in this estimator. This estimator deals with real signals having simulated anechoic array recording. This estimator shows better performance even under adverse conditions.

In 2012, Daniele Giacobello et al. [25] overviews the various linear predictors for speech analysis and coding. In this paper sparsity is introduced into the linear prediction framework. These sparse linear predictors perform more efficient decoupling between the pitch harmonics and the spectral envelope. This leads to the uncorrupted predictors of the pitch excitation. Properties such as shift invariance and pitch invariance are also offered by these predictors. A more synergistic new approach is proposed in this paper to encode a speech segment along with a compact representation. This approach also reduces the size and cost of the computations required.

In 2013, Jasper Kjaer Nilsen et al [26] used both a real- and complex-valued periodic signals along with additive noise as input in order to derive the probability model. In this paper the prior information is turned intoobservation model and prior distributions which are further turned into g-prior which is the more convenient prior. This is done using approximation on signal-to-noise-ratio (SNR) and the number of observations. In this paper the posterior

distribution is also derived from the fundamental frequency. Comparison between various approximations is also done in this paper. Result of this comparison concluded that the BIC approximation is worse than the other approximations.

In 2013 Jasper Kjaer Nilsen et al. [27] both the DOA and the pitch of harmonic source is estimated using ULA. This technique reveals that if harmonic structure is taken into account, the estimation done for DOA is more accurate. Also, the use of multiple sensors increases the accuracy of pitch estimation. Two estimators named as NLS and aNLS are proposed in this paper. This method gives better results as compared to the other methods in terms of mean squared error. Performance is also enhanced through this method and this method is also applicable on real-life signals.

In 2014 Zhiyao Duan et al. [28] Constrained-clustering approach is proposed. This approach works on harmonic sound sources ad perform the streaming of multiple pitches. Estimation of pitches is done in time frames using multi-pitch estimation (MPE) algorithm. No pre-training is required for systems following this approach and this approach is applicable to both music and speech signals. A new cepstrum named as uniform discrete cepstrum (UDC) is proposed, which presents the timbre of sound sources.

IV. SUMMARIZATION

S. No.	Year	Name of Author	Algorithms Used
1	1999	P. J. Walmsley et al [4]	Polyphonic pitch tracking using joint Bayesian estimation
2	2001	A. P. Klapuri [7]	spectral smoothness principle
3	2003	A. P. Klapuri [9]	Based on Harmonicity and Spectral Smoothness
4	2004	S. S. Abeysekera [10]	Using frequency-lag domain and BiSpectrum
5	2008	X. Zhang et al [12]	Based on Weighted Summary Correlogram
6	2008	A. Klapuri [13]	Using an Auditory Model
7	2009	R. Bedeau et al [14]	Expectation Maximization algorithm
8	2010	E. Vincent et al [15]	Adaptive Harmonic Spectral Decomposition
9	2011	E. Benelos et al [16]	Using Harmonic Envelope Estimation
10	2011	A Koretz et al [17]	Based on Maximum A Posteriori Probability
11	2011	Q Huang et al [18]	Based on Multi-Length Windows Harmonic Model
12	2011	Wohlmayr et al.[20]	Using Factorial Hidden Markov approach
13	2011	Mads Graesboll et al.[23]	Model order using posteriori Principles
14	2012	John Xi Zhang et. al.[24]	Subspace analysis along with time Space Model
15	2012	Daniele Giacobello et. al. [25]	Combination of Sparsity and linear prediction framework
16	2013	S. I. Adalbjornsson [19]	Using block sparsity
17	2013	Jasper Kjaer Nilsen et. al. [26]	Prior information turned into observation model and prior distribution into g-prior . Result come with probability Model.
18	2013	Jasper Kjaer Nilsen et. al. [27]	DOA and the pitch of harmonic source is estimated using ULA
19	2014	Zhiyao Duan et. al.[28]	Constrained Clustering approach

Table 1: Summarization of multi pitch detection algorithms in speech.

V. CONCLUSION

This paper concludes with evaluation of various algorithms being used in order to detect multiple pitches found in speech signal. Although it is difficult to detect exact 100% accurate details of speakers, the research is still being carried out. However, accuracy constraints are more with detection of number of speeches in one signal, tonal language and other dialect problems, gender identification within varying bit rate multi pitch detection etc. However, these can be implemented further for various applications like those of parliamentary debates, vocal discussion, telephonic meeting etc..

VI. REFERENCES

[1] A. M. Noll, "Short-Time Spectrum and Cepstrum Techniques for Vocal-Pitch Detection", *J.A.S.A.*, vol. 36, no. 2, February **1964**.

[2] A. M. Noll, "Cepstrum Pitch Determination", J.A.S.A., vol. 41, no. 2, **1967**.

[3] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg and C. A. McGonegal, "A Comparative Performance Study of Several Pitch Detection Algorithms", ASSP, vol. 24, no. 5, October **1976**.

[4] P. J. Walmsley, S. J. Godsill and P. J. W. Rayner, "Polyphonic pitch tracking using joint Bayesian estimation of multiple frame parameters", *Proc. 1999 IfiM Workshop on Applications & Signal Processing 10 Audio and Acoustics*, **1999**.

[5] M. Karjalainen and T. Tolonen, "Multi Pitch and periodicity analysis model for sound separation and auditory scene", *IEEE*, **1999**.

[6] M. Karjalainen and T. Tolonen, "A Computationally Efficient Multipitch Analysis Model", IEEE Transaction on Speech and Audio Processing, vol. 8, no. 6, November **2000**.

[7] A. P. Klapuri, "Multipitch extimation and sound seperation by the spectral smoothness principle", IEEE, **2001**.

[8] M Wu, D Wang and G. J. Brown, "A Multipitch Tracking Algorithm for Noisy Speech", IEEE Transaction on Speech and Audio Processing, vol. 11, no. 3, November **2003**

[9] A. P. Klapuri, "Multiple Fundamental Frequency Estimation Based on Harmonicity and Spectral Smoothness", IEEE Transaction on Speech and Audio Processing, vol. 11, no. 3, November **2003**.

[10] S. S. Abeysekera, "Multiple Pitch estimation of poly-phonic audio signals in a frequency-lag domain using the bispectrum", *IEEE*, **2004**.

[11] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "The multi pitch estimation problem: Some New Solution", *IEEE*, **2007.**

[12] X. Zhang, W. Liu, P. Li and B. Xu, "Multipitch Detection Based on Weighted Summary Correlogram", National Laboratory of Pattern Recognition, Beijing.

[13] A. Klapuri, "Multipitch Analysis of Polyphonic Music and Speech Signals Using an Auditory Model", IEEE Transaction on Speech and Audio Processing, vol. 16, no. 2, February **2008**.

[14] R. Badeau, V. Emiya and B. David, "Expectation Maximization algorithm for multi pitch estimation and separation of overlapping harmonic spectra", IEEE, **2009**.

[15] E. Vincent, N. Bertin and R. Badeau ,"Adaptive Harmonic Spectral Decomposition for Multiple Pitch Estimation", IEEE Transaction on Speech and Audio Processing, vol. 18, no. 3, March **2010.**

[16] E. Benetos and S. Dixon, "Joint Multi-Pitch Detection Using Harmonic Envelope Estimation for Polyphonic Music Transcription", IEEE Journal of selected topic in Signal Processing, vol. 5, no. 6, **2011.**

[17] A. Koretz and J Tabrikian, "Maximum A Posteriori Probability Multiple-Pitch Tracking Using the Harmonic Model", IEEE Transaction on Speech and Audio Processing, vol. 19, no. 7, September **2011**.

[18] Q Huang and D Wang, "Multi-Pitch Estimation for Speech Mixture Based on Multi-Length Windows Harmonic Model", Proc. Of IJCCSO, **2011**.

[19] S. I. Adalbjornsson, A. Jakobsson, and M. G. Christensen, "Estimating multiple pitches using block sparsity", *IEEE*, **2013**.

[20] M. Wohlmayr, M. Stark, and F. Pernkopf, "A probabilistic interaction model for multipitch tracking with factorial hidden Markov models," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 4, pp. 799–810, May **2011**.

[21] M. Bay, A. F. Ehmann, J. W. Beauchamp, P. Smaragdis, and J. S. Downie, "Second fiddle is important too: Pitch tracking individual voices in polyphonic music," in *Proc. Int. Soc. Music Inf. Retrieval Conf. (ISMIR)*, pp. 319–324, **2012**.

[22] E. Vincent, "Musical source separation using time-frequency source priors," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 1, pp. 91–98, **2006.**

[23] Mads Græsbøll Christensen, Jesper Lisby Højvang, Andreas Jakobsson and Søren Holdt Jensen," Joint fundamental frequency and order estimation using optimal filtering", Christensen et al. EURASIP Journal on Advances in Signal Processing **2011**.

[24] Johan Xi Zhang1, Mads Græsbøll Christensen, Søren Holdt Jensen and Marc Moonen," Joint DOA and multi-pitch estimation based on subspace techniques" EURASIP Journal on Advances in Signal Processing **2012**.

[25] Daniele Giacobello, , Mads Græsbøll Christensen, Manohar N. Murthi, , Søren Holdt Jensen, , and Marc Moonen, "Sparse Linear Prediction And Its Applications To Speech Processing", Ieee Transactions On Audio, Speech, and Language Processing, Vol. 20, No. 5, July **2012**.

[26] Jesper Kjær Nielsen, Mads Græsbøll Christensen And Søren Holdt Jensen , "Default Bayesian Estimation of The Fundamental Frequency" , Ieee Transactions On Audio, Speech, and Language Processing, Vol. 21, No. 3, March **2013.**

[27] Jesper Rindom Jensen, Mads Græsbøll Christensen And Søren Holdt Jensen "Nonlinear Least Squares Methods For Joint Doa And Pitch Estimation", *Ieee Transactions On Audio, Speech, And Language Processing*, Vol. 21, No. 5, May **2013**.

[28] Zhiyao Duan, Jinyu Han, and Bryan Pardo*," Multi-Pitch Streaming Of Harmonic Sound Mixtures ", Ieee/Acm Transactions On Audio, Speech, And Language Processing, Vol. 22, No. 1, January **2014.***

