IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AUDIO SOURCE SEPARATION AND NOTE PROCESSING AND INSTRUMENT DISPLAY

[1]Minnuja Shelly, [2]Gopika T G, [3]Safrin Sukkur, [4]Shreethal Janardhanan, [5]Shuhaib P M

[1]Guide, [2]Student, , [3]Student, [4]Student, [5]Student
[1]Dept. of Computer Science And Engineering,
[1]Universal Engineering College, Thrissur, India

Abstract: For the past few years ,we have seen many applications related to playing and teaching musical instruments. But most of the apps have their set of songs to play and also some of them may not involve our own choice which is a major limitation of these applications. This proposed system solves this both issues. The main intention of this system is for the user to enjoy the time of usage. In this system, audio as input is fed into the system where piano notes separated from the in put and these notes are highlighted on the screen piano key display along with the music. Many systems have already been developed which generate sheet music by taking song as input. This paper presents one such system that aims to display the corresponding key along with the song with an emission of key as per the piano notes. Improvement is achieved by working on source separation and chord estimation modules of the previous system. There are huge number of artists already present globally in the music industry. Apart from them, music professionals, students, learners and amateurs also form part of this large music community. Virtuo piano is one of the sources for learning and playing songs for these artists and amateurs and also this can be used even by a duff person. Virtuo piano is an aid in the form of a application consisting of keys on display and sheet music with chords lyrics of a song, and is very helpful to piano even for a duff person. It enables players to identify the corresponding key, chords, rhythm, pitch to any song.

***Index Terms* - Component, formatting, style, styling, insert.**

I. INTRODUCTION

Virtuo piano is a piano learning app and Recurrent Neural Network is used for the music application. There are a large number of artists already in the music industry around the world. Apart from them, music professionals, students, and amateurs are also part of this great music community. The proposed system is a sources for learning and playing the songs with a guidance of emitting key on piano key board in app display .And sheet music also displayed on the screen. In the past, there had been several apps that used definite song sets which could only be used on a specific set of music. But in this proposed system it can separate song which given as input to the system. It enables players to identify piano, vocal, bass and drums in any song gives in input. The underlying system use RNN and ISTM to identify the piano and other elements need to be separated. The primary goal music source separation and audio post-processing is done by the MIDI (MSSAPPMIDI).

From many years, sheet music was the source has been used by musicians to practice music and pass it on to future generations as well. As everything has fallacies, sheet music has some problems too. The first problem is unavailability of sheet music and it is one of the major problems faced by artists presently and this unavailability may be due to various reasons including copyright issues where access to any resource is denied to others through formal rules and laws. Secondly, some local songs or songs that are not so famous may not have music sheet present at all. There may be a situation where a music artist, amateur or student, with not much experience, may want to play such a song on a piano but doesn't know the chords. In such cases, these artists and amateurs are left helpless without any another source since amateurs heavily rely on sheet music to fulfill their ambition of playing others' song. A student can apply his/her learning through practice with the help of virto piano by learning with the emitting key display which is emits corresponding to the particular song's piano note. Practice is possible by music sheet display too. If the user is awed about sheet music then he/she can play by the sheet display . if the user is a fresher to piano playing then they can use the help of key-board display which emits key corresponding to the piano notes along the song.

II. LITERATURE SURVEY

Here we introduce each papers based on the technologies used in the Audio Source Separation And Note Processing Along With Education and they are arranged in technologies bases.

Taiwanese researchers have developed a sheet music generator which is based on deep learning techniques. The proposed system would be a great help if sheet music can be automatically generated for all kinds of music lovers, such as professional musicians, music enthusiasts, amateurs, and other potential users. The DeepSheet system needs to be equipped the capability to detect the pauses and chords from a given audio wave data.

For the problem of separating concur- rent speech through a spatial filtering stage and a subsequent time-frequency masking stage. Log-sigmoid masks are optimized to maximize the kurtosis of the separated speech. Experiments on the Pascal Speech Separation Challenge II show significant improvements compared to previous approaches.

Using the kurtosis as a scale parameter for speech separation, we investigate the use of a linear constrained minimum variance (LCMV) beam former. This gave significantly better results than a super directive beam former and was only slightly inferior to the MMI beam former[1].

The proposed system first applies a.ndouble stage HPSS as pre-processing. Features are then extracted from a filter bank on a Mel scale and supplied as input to the deep. The blocks of our system are described in more de- tails below. We are working on a deep neural network that learns how to classify binary signals. The architecture of the network is determined by the hidden layers and the number of neurons or LSTM blocks within each layer.

Chords are defined by the simultaneous sounding of two or more musi- cal notes. The interval relationships between these notes determine the type of chord. We aim to design a system capable of classifying audio frames as one of 108 chords, including 12 variations of major, minor, augmented, suspended 2nd and suspended 4th. Some systems make use of the constant-Q transform [1, 6, 8] with some of these employing a tuning algorithm to allow for differences in tuning. Other approaches calculate the chromagram directly from the dis- crete Fourier transform (DFT) of the input signal.Some existing chroma calculation techniques include all energy within a given fre- quency range in the amplitude value of a certain pitch class. Our approach identifies only the energy in the harmonics within a given range. The square root of the magni- tude spectrum is taken to reduce the amplitude difference between harmonic peaks. The chroma vector,C, is calculated by.P=12, the number of notes playing a rhythmic accompaniment, and. L(f) where L is the frame size and F is the sampling frequency.

Our approach ex- amines 2 octaves of the spectrum, between 130.81Hz and 523.25 Hz. We hypothesize that the majority of instruments playing a rhythmical accompaniment play the lower register of the instrument. Each layer contains one unique neuron with a logistic activation function, whose output is an estimated probability of presence. In [20] we show that in a super-vised gradient-trained deep neural network with random weights initialization, layers far from the output are poorly optimized. In our work, we extended this procedure to automati- cally learn the network architecture during the training. We used the Jamendo Corpus, a publicly avail-able dataset including singing voice activity annotations. Over-fitting is controlled by early-stopping: training starts with a step for the gradient descentη= 10 −5 and a momentum= 0.9. Sudo RM -RF: Efficient networks for universal audio source separation is easy to see that the proposed models can match and even out-perform the separation performance of other several state-of-the-art models using orders of magnitude less computational requirements.

On par with many state-of-the-art approaches in the literature [2, 9, 3, 6], SuDoRM-RF performs end-to-end audio source separation using a mask-based architecture with adaptive encoder and decoder basis. The input is the raw signal from a mixture x ∈ R T with T samples in the time-domain. First we feed the input mixture x to an encoder E in order to obtain a latent representation for the mixture vx = E (x) ∈ R CE×L Consequently the latent mixture repre- sentation is fed through a separation module S which estimates the corresponding masks mb i ∈ R CE×L for each one of the N sources s1, · · · , sN ∈ R T which constitute in the mixture. The estimated la- tent representation for each source in the latent space vbi is retrieved by multiplying element-wise an estimated mask mb i with the en- coded mixture representation vx. Finally, the reconstruction for each source bsi is obtained by using a decoder D to transform the latent- space vbi source estimates back into the time-domain bsi = D (vbi). An overview of the SuDoRM-RF architecture is displayed in Fig- ure 1. The encoder, separator and decoder modules are described in Sections 2.1, 2.2 and 2.3, respectively. For simplicity of our notation we will describe the whole architecture assuming that the processed batch size is one. Moreover, we are going to define some useful operators of the various convolutions which are used in SuDoRM-RF.

The object detector is trained for a vocabulary of C objects. In general, this detector should cover any potential soundmaking object categories that may appear in training videos. The implementation uses the Faster R-CNN [36] object detector with a ResNet-101 [17] backbone trained with Open Images [26]. For each unla- beled training video, uses the pre-trained object detector to automatically find objects in all video frames. Then, gather all object detections across frames to obtain a video-level pool of objects. For Audio-Visual Separator the detected object re- gions to guide the source separation process. A related design for multi-modal feature fusion is also used in [13, 30, 31] for audio spatialization and separation. How- ever, unlike those models, our separator network combines the visual features of a localized object region and the audio features of the mixed audio to predict a magnitude spectro- gram mask for source separation. The network takes a detected object region and the mixed audio signal as input, and separates the portion of the sound responsible for the object. A ResNet-18 network is used to extract visual features after the 4th ResNet block with size (H/32) × (W/32) × D, where H, W, D denote the frame and channel dimensions. Then pass the visual feature.

III. THE PREPOSED SYSTEM METHODOLOGY

In this proposed system includes Artificial intelligence and RNN(Recurrent Neural Network) technologies. The proposed system music and vocals are separated from audio and the bass source and piano source are extracted from it . the frequency of the particular music and piano keys are compared for match and highlights the key which the frequency is matched with a yellow color .

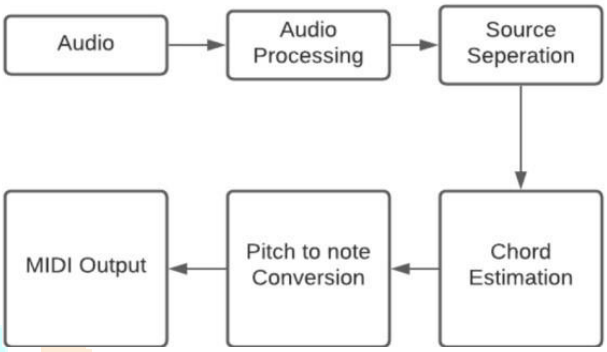


fig.Block diagram

3. 1 Phases Of The System

The methodology section outline the plan and method that how the study is conducted. This includes Universe of the study, sample of the study, Data and Sources of Data, study's variables and analytical framework. The details are as follows

3.1.1 Training phase

In this phase the proposed system trains the MIDI files (downloaded from classical piano midi page) using LSTM.

3.1.2 Noise removing phase

Then uploading the audio and processing by removing noise and unwanted parts. Then connects with MIDI files.

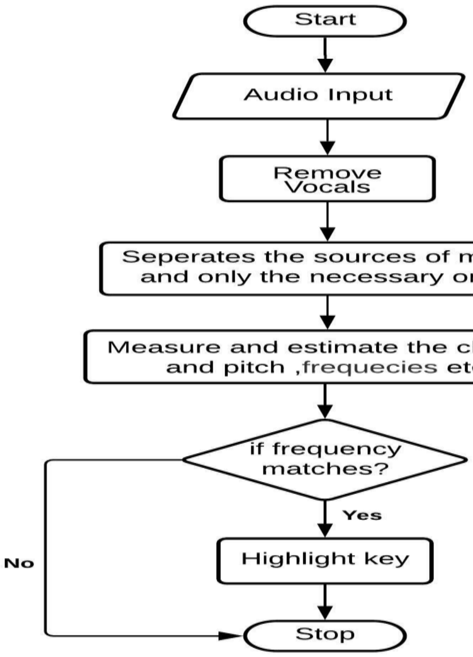
3.1.3 Background separation phase

In this phase the proposed system identifies piano using LSTM and RNN by comparing the MIDI notes and separates it . The separated notes will stored in a created folder called "Piano.mp3". In this phase it also generates and separates the drums , vocals and bass for this purpose the system uses a package called "splitter".

3.1.4 User phase

In this phase the proposed system connects the user inter phase with the drive. Converts MIDI files in to notes by measuring frequency and pitch and emits the corresponding key with a yellow color.

3.2 Activity diagram:


Start
Audio Input
Remove Vocals
Seperates the sources of music and only the necessary ones
Measure and estimate the chords and pitch ,frequecies etc
if frequency matches?
Yes
No
Highlight key
Stop

Here we introduce each papers based on the technologies used in the Audio Source Separation And Note Processing Along With Education and theyare arranged in technologies bases.

Taiwanese researchers have developed a sheet music generator which is based on deep learning techniques. The proposed system would be a great help if sheet music can be automatically generated for all kinds of music lovers, such as professional musicians, music enthusiasts, amateurs, and other

potential users. The DeepSheet system needs to be equipped the capability to detect the pauses and chords from a given audio wave data.

For the problem of separating concur- rent speech through a spatial filtering stage and a subsequent time-frequency masking stage. Log-sigmoid masksare optimized to maximize the kurtosis of the separated speech. Experiments on the Pascal Speech Separation Challenge II show significant improvements compared to previous approaches.

Using the kurtosis as a scale parameter for speech separation, we investigate the use of a linear constrained minimum variance (LCMV) beam former. This gave significantly better results than a super directive beam former and was only slightly inferior to the MMI beam former[1].

The proposed system first applies a double stageHPSS as pre-processing. Features are then extracted from a filter bank on a Mel scale and supplied asinput to the deep. The blocks of our system are described in more de- tails below. We are working ona deep neural network that learns how to classifybinary signals. The architecture of the network is determined by the hidden layers and the number of neurons or LSTM blocks within each layer.

Chords are defined by the simultaneous sounding of two or more musical notes. The interval relationships between these notes determine the type of chord. We aim to design a system capable of classifying audio frames as one of 108 chords, including 12 variations of major, minor, augmented, suspended 2nd and suspended 4th. Some systems make use of the constant-Q transform [1, 6, 8] with some of these employing a tuning algorithm to allow for differences in tuning. Other approaches calculate the chromagram directly from the discrete Fourier transform (DFT) of the input signal. Some existing chroma calculation techniques include all energy within a given fre- quency range in the amplitude value of a certain pitch class. Our approach identifies only the energy in the harmonics within a given range. The square root of the magnitude spectrum istaken to reduce the amplitude difference between harmonic peaks. The chroma vector,C, is calculated by.P=12, the number of notes playing a rhythmicaccompaniment, and. L(f) where L is the frame size and F is the sampling frequency.

Our approach ex- amines 2 octaves of the spectrum, between 130.81Hz and523.25Hz. We hypothesise that the majority of instruments playing a rhythmical accompaniment play the lower register of the instrument. Each layer contains one unique neuron with a logistic activation function, whose output is anestimated probability of presence. In [20] we show that in a super- vised gradient-trained deep neural network with random weights initialization, layers farfrom the output are poorly optimized. In our work, we extended this procedure to automati- cally learn the network architecture during the training. We used the Jamendo Corpus, a publicly avail-able dataset including singing voice activity annotations. Over- fitting is controlled by early-stopping: training starts with a step for the gradient descentη= 10 −5 and a momentum$m$= 0.9. Sudo RM -RF: Efficent networks for universal audio sorce separation is easy to see that the proposed models can match and even out-perform the separation performance of other several state-of-the-art models using orders of magnitude less computational requirements.

On par with many state-of-the-art approaches in the literature [2, 9, 3, 6], SuDoRM-RF performs end-to- end audio source separation using a mask-based architecture with adaptive encoder and decoder basis.The input is the raw signal from a mixture $x \in R$ T with T samples in the time-domain. First we feed the input mixture x to an encoder E in order to obtain a latent representation for the mixture vx = E (x) ∈ R CE×L Consequently the latent mixture repre- sentation is fed through a separation module S which estimates the corresponding masks mb i ∈ R CE×L for each one of the N sources s1, · · · , sN ∈ R T which constitute in the mixture. The estimated la- tent representation for each source in the latent space vbi is retrieved by multiplying element-wise an estimated mask mb i with the en- coded mixture representation vx. Finally, the reconstruction for eachsource bsi is obtained by using a decoder D to transform the latent- space vbi source estimates back into the time-domain bsi = D (vbi). An overview of the SuDoRM-RF architecture is displayed in Fig- ure1. The encoder, separator and decoder modules are described in Sections 2.1, 2.2 and 2.3, respectively. For simplicity of our notation we will describe the whole architecture assuming that the processed batch size is one. Moreover, we are going to define some useful operators of the various convolutions whichare used in SuDoRM-RF.

The object detector is trained for a vocabulary of C objects. In general, this detector should cover any potential soundmaking object categories that may appear in training videos. The implementation uses the Faster R-CNN [36] object detector with a ResNet-101 [17] backbone trained with Open Images [26]. For each unla- beled training video, uses the pre-trained object detector to automatically2 find objects in all video frames. Then, gather all object detections across frames to obtain a video-level pool of objects. For Audio-Visual Separator the detected object regions to guide the source separation process. A related design for multi-modal feature fusion is also used in [13, 30, 31] for audio spatialization and separation. How- ever, unlike those models, our separator network combines the visual features of a localized object region and the audio features of the mixed audio to predict a magnitude spectro- gram mask for source separation. Thenetwork takes a detected object region and the mixed audio signal as input, and separates the portion of the sound responsible for the object. A ResNet-18network is used to extract visual features after the 4thResNet block with size (H/32) × (W/32) × D, where H, W, D denote the frame and channel dimensions. Then pass the visual feature.

IV. CONCLUSION AND FUTURE SCOPE

The proposed system is a simple demonstration of implementation of deep learning using RNN with better accuracy. The system helps in source separation and generation of piano music. Unlike the traditional system where there was a defined dataset for piano instrument aspirants , in the proposed system any song can be applied as input.

Currently the proposed system only works for keyboard (piano) . In future it can be extended to work for ather instruments like guitar, harp etc.with enough data set and training . The improved application can also be used to assess the accuracy when played in a real keyboard .

IV. ACKNOWLEDGMENT

We wish to record our indebtedness and thankfulness to all those who helped us to prepare this report titled “VIRTUO PIANO” and present it in a satisfactory way.

First and foremost, we thank God Almighty for His providence and for being the guiding light throughout the seminar.

We feel highly privileged in making a mention of Dr.Jose K Jacob -Principal, UEC for his co-operation and help. We deeply and wholeheartedly thank Mr.Sreeraj R -HOD, CSE for his extreme valuable advice and encouragement.

We are especially thankful to our guide and supervisor Ms. Minnuja Shelly – Assistant Professor, UEC for giving us valuable suggestions and critical inputs in the preparation of this report. We would also like to extend our heartfelt thanks to our project coordinator Ms. Chinju Poulose- Assistant professor, UEC for her meticulous guidance and support that helped us in the completion of our project.

We would like to extend our sincere gratitude to all faculty of Computer Science and Engineering department for the support and suggestions that helped us in the development of ourproject to what it is now. We thank our parents for the mental support provided during the course of the project at the times when our energies were the lowest.

REFERENCES

[1] S.Chen,B.Mulgrew,andP.M.Grant,―A clustering technique for digital communications channel equalization using radial basis function networks, *IEEETrans.onNeural Networks*, vol.4,pp.570-578,July1993

[2] J. U. Duncombe, ―Infrared navigation—Part I: An assessment of feasibility, *IEEE Trans. Electron Devices*,vol.E D-11,pp. 34-39,Jan.1959.

[3] C.Y.Lin, M.Wu, J.A.Bloom, I.J.Cox, and M.Miller,―Rotation, scale, and translation resilient public watermarking for images,*IEEE Trans.Image Process.*,vol. 10,no.5,pp.767-782, May2001.

[4] A.Cichocki and R.Unbehaven, *Neural Networks for Optimization and Signal Processing*, 1sted. Chichest er,U.K.:Wiley, 1993,ch.2,pp.45-47.

[5] K.Chen, *Linear Networks and Systems*, Belmont, CA:Wadsworth, 1993, pp. 123-135.

[6] H.Poor, *An Introduction to Signal Detection and Estimation*; NewYork: Springer- Verlag, 1985,ch.4.

[7] BJ.Williams, ― Narrow-band analyzer, ‖ Ph.D .dissertation, Dept. of Electrical.Engineering., Harvard Univ. ,Cambridge, MA,1993.

[8] R.A.Scholtz,―The Spread Spectrum Concept, ‖in *Multiple Access*, N.Abramson, Ed. Piscataway, NJ: IEEE Press, 1993,ch.3, pp. 121-123.

[9] G.O.Young,―Syntheticstructureofindustrialplastic s,‖ in*Plastics*,2nded.vol.3,J.Peters,Ed.NewYork:Mc Graw-Hill,1964,pp.15-64.

[10] SW.D.Doyle,―Magnetization reversal in films with biaxial anisotropy,‖ in*Proc.1987 INTERMAG Conf.*, 1987,pp. 2.2-1-2.2-6

[11] G. W JuetteandL. E. Zeffanella, ―Radio noise currents n short sections on bundle conductors, presented at the IEEE Summer Power Meeting, Dallas, TX, June 22-27,1990.

[12] J.P.Wilkinson,―Nonlinear resonant circuit devices,‖ U.S.Patent3624 12,July16, 1990

[13] *Letter Symbols for Quantities*, ANSI StandardY10.5-1968.

[14] *Transmission Systems for Communications* ,3rded., Western Electric Co.,Winston-Salem, NC, 1985, pp. 44-60.

[15] *Motorola Semiconductor Data Manual,* Motorola Semi-conductor Products Inc., Phoenix, AZ, 1989.

[16] R.J.Vidmar.(August1992).On the use of atmospheric plasmas as electromagnetic reflectors.*IEEE Trans. Plasma Sci.* [Online]. 21(3).pp.876-880.Available: http://www.halcyon.com/pub/journals/21ps03-vidmar