Extraction of Audio Signal Features using Dynamic Mode Decomposition (Early Research Final Project)

Oluwafemi Olaleke [1]

Abstract

Dynamic Mode Decomposition is a numerical procedure for extracting dynamical features from flow data .It is originally introduced in the fluid mechanics community, and it has emerged as a powerful tool for analyzing the dynamics of nonlinear system. DMD relies only on the high-fidelity measurements, like experimental data and numerical simulations, so it is an equation-free algorithm. It is popularity is also due to the fact that it does not make any assumptions about the underlying system.As many DMD applications exist in other field but not in the area of sound data, my early research project focuses on examining the possible feature extraction outcomes that can be realized in applying this Algorithm on sound data and using these possible features for sound classification task.

1. DMD on Audio Signal:The Methodology

From my research on the vast application of DMD in various fields(Tu et al., 2013), I discovered that in other to effectively employ the numerical power of DMD to extracting useful features, certain crucial steps must be taken in terms of data arrangement. In achieving the core purpose of my task, the following procedures were followed specifically in applying DMD on audio signals.

1.1. Getting the Audio Signal into appropriate form for DMD operation

The DMD operation is based on a special form of data arrangement. A matrix much be formed from the flow data such that the columns of this matrix are time-shifted version of each other. An audio signal can be brought into such form through the following computations;

- An audio signal can be represented by a series of measurements y_t at constant time intervals $t = k\Delta t$
- Over a short time frame, the signal is formed by superimposing N simple oscillators with angular frequencies $w_{j=1}^{N}$ and complex amplitudes $A_{j=1}^{N}$
- Mathematically, such signal is formed as:

$$y_t = Re\left(\sum_{j=1}^{N} A_j exp(iw_j k\Delta t)\right)$$

- An audio signal vector z_y can be computed from the y_t result above such that;

$$z_t = [y_t......y_{t+2N-1}]$$

  Its important to delay observations with at least $2N$ in order to detect N frequencies (Kamb et al., 2020)

- The next step is to arrange z_t and its time-shifted version as columns of two bigger matrices, and this I did using inspiration the Henkel matrix construction, such that:

$$Z_t = [z_t......z_{t+N-1}]$$

$$Z_{t+1} = [z_{t+1}......z_{t+N}]$$

- The final step for the data re-arrangement is to compute the Matrix K_t for DMD operation as:

$$K_t = Z_{t+1} Z_t^{+}$$

The matrix K_t gives an appropriate matrix for the DMD operation.

1.2. DMD operation: Audio Frequency from DMD Eigenvalues

Generally, DMD analyzes the dynamics of nonlinear system, it gives information about the mode and dynamics of the system being considered. Different variation of DMD has been developed due to vast applications of this method.

For my sound data experiment , I applied Higher Order-DMD (**?**) higher which is basically applied on 1D snapshots, and the resulting Eigenvalues, Dynamics and Mode from this operation were evaluated to get the needed sound features in the following way;

- I discovered it was possible to compute Sound Frequency from the resulting Eigenvalues of DMD. this is because, the eigenvalues λ_j correspond to the temporal dynamics of each spatial mode ϕ_j (Brunton et al., 2016)
- Specifically the rate of growth/decay of the mode and frequency of oscillation are reflected in the magnitude and phase components of the eigenvalues
- While the magnitude of the eigenvalues relative to the unit circle indicates whether the corresponding mode is growing or decaying, the phase of each eigenvalue translates to the frequency of oscillation. and its computed as;

$$f_j = \frac{1}{2\pi\Delta t} Imag(log\lambda_j)$$

1.3. Computing DMD Power spectrum from spatial Modes

The magnitude of each mode represents spatial correlations between the observable n locations,and the power spectrum can be computed from these modes;such that

$$DMDPowerspectrum = ||Mode||^2$$

2. Mel Frequency Cepstral Coefficients (MFCCs)

After successfully extracting the audio signal frequencies and Spectrum, the next step is to compute the Mel Frequency Cepstral Coefficients (MFCC) from the DMD power spectrum

In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.

The MFCCs is good at identifying the linguistic components of an audio signal, because it accurately determines the shape of the vocal tract by which the signal is produced. This gives us the representation of the phoneme being produced. The shape of the vocal tract manifests itself in the envelope of the short time power spectrum, and MFCCs accurately represent this envelope.

Since the MFCC summarises the frequency distribution across the window size, so it is possible to analyse both the frequency and time characteristics of the sound. These audio representations will allow me to identify features for classification.

I followed the following steps in computing MFCC from the pre-computed DMD spectral.

- Energy Evaluation from DMD-Spectral : I evaluated the energy corresponding to each of the DMD spectrum. I did this by applying a series of mel filterbanks to the power spectra. This gave me two output; The first is a matrix containing the Mel-filterbank energies of size (num-frames by num-modes),and the second is the energy values of each frame
- MFSC computation from filter energies: From the filter energies in the previous step, I computed the MFSCs, by first taking the logarithm of the mel-filter bank energy. Using logarithm here allows the use of the cepstral mean subtraction, which is a channel normalisation technique. The logarithm of the filter-Bank energy gives the *Mel Frequency Spectral Coefficient[MFSC]*. Then by computing the Discrete Cosine Transform(DCT) of the log filterBank energy(MFSC), I was able to get the Mel fequency Ceptral Coefficient(MFCC).

However, I discovered there is a simpler way of getting the MFCC values by relying on *librosa functions*. The librosa function for audio processing has already make it easy to extract the MFCC from a pre-computed Power Spectrum. I inputted the DMD Spectral values into the *librosa.feature.melspectrogram* to get the Mel-Spectrogram values of the audio signal.

By using the this Mel-Spectrogram values, I computed MFCCs using *librosa.feature.mfcc*

3. Sound Classification Task with Jax and Haiku

One of the goals of this project is to employ the features extracted using DMD into Audio recognition task by performing a Sound Classification task. I studied about building Neural Network based models in an easy way with Haiku and Jax, and I made an initial attempt using a classic sound classification method on the ESC-50 dataset.

The dataset is a labeled collection of 2000 environmental audio recordings suitable for benchmarking methods of environmental sound classification.It consists 5-second-long recordings organized into 50 semantical classes (with 40 examples per class) loosely arranged into 5 major categories. The Sound waves are digitised by sampling them at discrete intervals known as the sampling rate (typically 44.1kHz for CD quality audio meaning samples are taken 44,100 times per second).

For this first attempt of my experiment, I used MFCCs gotten using the librosa function based on the Short Time Fourier Transform of audio signals.

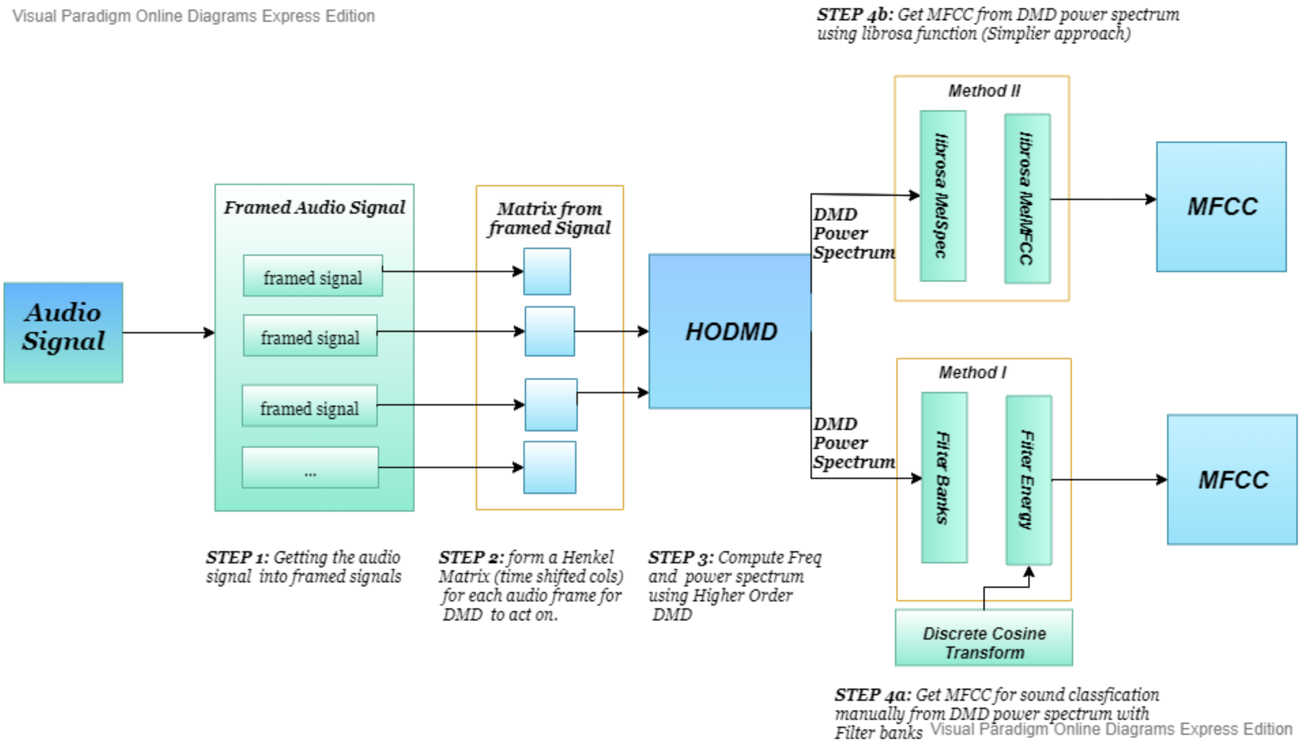
Visual Paradigm Online Diagrams Express Edition
STEP 4b: Get MFCC from DMD power spectrum using librosa function (Simplier approach)
Method II
librosa MelSpec
librosa MelMFCC
MFCC
Audio Signal
Framed Audio Signal
framed signal
framed signal
framed signal
...
Matrix from framed Signal
HODMD
DMD Power Spectrum
DMD Power Spectrum
Method I
Filter Banks
Filter Energy
MFCC
Discrete Cosine Transform
STEP 1: Getting the audio signal into framed signals
STEP 2: form a Henkel Matrix (time shifted cols) for each audio frame for DMD to act on.
STEP 3: Compute Freq and power spectrum using Higher Order DMD
STEP 4a: Get MFCC for sound classfication manually from DMD power spectrum with Filter banks
Visual Paradigm Online Diagrams Express Edition

Figure 1. Block Diagram summarizing steps taking in extracting MFCCs using DMD. Step 2 might be discarded as HODMD handles data re-arrangement for simplier computation

3.1. Multi Layer Perceptron - (MLP) based Model

I built an MLP based model for sound classification by bringing together the necessary functions with Jax; starting with a function that initialzes the weights of the Neural Network, and returns a list of layer-specific parameters, and other that perform forward pass through the network,that computes the cross-entropy loss of the predictions (classification task),evaluation of accuracy of the prediction and another that updates the parameters using some form of gradient descent.

I used the extracted MFCCs based on STFS as my feature input and trained appropriately.

3.2. Convolutional Neural Network (CNN) based model

For the CNN based model, I made a modification in the feature extraction stage, because CNNs required a fixed size for all inputs. This modification was done by zero padding the output vectors to make them all the same size. The CNN model is made up of three layers and I included maxpooling ,batch normalizing and dropout to avoid overfitting.

Link to my sound classification task using MFFcs from STFT [1].

3.3. Classification using MFCCs extracted via DMD Spectral

I attempted using the MFCCs gotten via my DMD spectral for this classification purpose. However, most of the ESC50 dataset gave a serious challenge during the DMD computing. This was because some frames of the sound datset had long silent regions which made an error to occur as the zero values in the frame made the DMD operation impossible in those frames.

Therefore I could not utilize all the full dataset for my experiment. Audio samples without silent region gave very good output.

Link to DMD extration processes and MFCCs computations [2].

[1] https://colab.research.google.com/drive/12wdbUR2wjGfttcUh2NMQ9A2Zw8ugafDM?usp=sharing

[2] https://colab.research.google.com/drive/1KnO85B5m_sbfcUIc1NoTjz9DccFcE968?usp=sharing

4. Conclusion

During this research project, I explored the possibility of extracting meaningful features from audio dataset using Dynamic Mode Decomposition; a data driven method that is widely used in the field of fluid mechanics.

I was able to successfully get meaning features from audio signals by exploring the conceptual interpretation of the eigenvalues and model outputs from Higher Order DMD output.

I also successfully computed Mel Frequency Ceptral Coefficients of audio signal from pre-computed DMD spectral.

I experimented Automatic Speech Recognition task by performing a sound classification task using Neural Network based models built using Jax and Haiku, first by using the classical MFCCs from Short Time Fourier Transform and then by using MFCCs obatained from DMD spectral.

References

Brunton, B. W., Johnson, L. A., Ojemann, J. G., and Kutz, J. N. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of neuroscience methods*, 258:1–15, 2016.

Kamb, M., Kaiser, E., Brunton, S. L., and Kutz, J. N. Time-delay observables for koopman: Theory and applications. *SIAM Journal on Applied Dynamical Systems*, 19(2):886–917, 2020.

Tu, J. H., Rowley, C. W., Luchtenburg, D. M., Brunton, S. L., and Kutz, J. N. On dynamic mode decomposition: Theory and applications. *arXiv preprint arXiv:1312.0041*, 2013.

Skoltech

Skolkovo Institute of Science and Technology

Early Research report form

Please submit this form to the Canvas assignment in your Thesis course (MSc Thesis Class 2021: MSc program) by 5 June 2020.
It is allowed to submit your report in free form if your Research Advisor does not mind.

Student's first name OLUWAFEMI

Student's last name OLALEKE

Date of report submission (e.g. 25 May 2020) 5th June 2020

Project supervisor's full name

Professor Ivan Oseledets

Project supervisor's work place and position

Project name in English (mandatory)

Extraction of Features from Audio Signal with Dynamic Mode Decomposition

Project name in Russian (mandatory)

Извлечение признаков из аудиосигнала с помощью разложения по динамическим модам

Project description

- This project involves investigating the possibility of getting meaningful features from Audio signals using Dynamic Mode Decomposition.

- Dynamic mode decomposition is a numerical procedure for extracting dynamical features from flow data. It is originally introduced in the fluid mechanics community, and it has emerged as a powerful tool for analyzing the dynamics of nonlinear system.

- DMD relies only on the high-fidelity measurements, like experimental data and numerical simulations, so it is an equation-free algorithm. It is popularity is also due to the fact that it does not make any assumptions about the underlying system.

Project goals and your personal goals in the projects. Describe your role in the project

- The main project goal is to effectively apply DMD on sound data, by first getting the sound data into a form DMD can act on,then getting useful features by running DMD experiments on the modified sound data sample.

- The resulting sound features gotten via DMD is then aimed to be tested for sound classification purposes.

Describe what you have accomplished (be specific) and the progress made toward the objective/goal

- I was success in understanding the DMD Algorithm, how it works and why it works, and especially its feature extraction applications in various field

- I was able to successfully understand the structure of sound data and I applied Higher Order DMD to extract useful features(frequency , Spectrum) needed for Audio Signal Recognition

- I learnt about JAX and HAIKU Neural Network based models,for sound classification

Describe any significant problems you faced and solutions found. If a project took a different course and your tasks/goals changed (from initially planned), describe why

- I had the problem of slow progress due to my computer being low standard.

-

Summarize the outcomes of the project for you (skills/competences developed, describe what you learned about working in a team)

- Successful extraction of Frequency and Power Spectrum from audio signal using DMD

- Extraction of Mel Frequency Ceptral Coefficients (MFCCs) from DMD power spectrum above

- Sound classification experiment using JAX and HAIKU.

Identify the format of the report to the project supervisor (presentation, detailed report, etc.)

Detailed Report

Student's confirmation:

☑ **Hereby I confirm that the report is approved by the project supervisor and academic or research (thesis) advisor (if known).**