Music Information Retrieval
Fall 2020

Midterm Exam
Time Limit: 100 minutes

Answer the following questions. Submit your answers in a single *Jupyter Notebook* (**.ipynb**) file. Mark your answers clearly and include any code you used. You may consult reference material, but references *must* be cited in your response. Do not consult with classmates. Responses must be your own, original work. This exam has 24 questions and is worth 20 points.

True or False

(0.25 points each)

1. A signal $x[n]$ is periodic when $x[n] = x[n + p]$ for some p.
2. The Discrete Fourier Transform of a real-valued signal is symmetric.
3. In supervised classification, the validation set is used to select hyperparameters of the model.
4. The first coefficient of Mel Frequency Cepstral Coefficients measures the overall loudness of an audio signal.
5. The Short Time Fourier Transform and the Spectrogram are not equivalent.
6. The input of an energy-based novelty function is the STFT of an audio signal.
7. Loudness is directly measurable, rather than a perceived quantity.
8. A sinusoid is not fully described by its frequency and amplitude.

Definitions

For each of the following, write a 1-2 sentence definition. (0.5 points each)

9. window function
10. log-frequency spectrogram
11. spectral centroid
12. feature vector
13. f-measure
14. timbre

Short Answer

(0.75 points each)

15. MIDI note 100 corresponds to what center frequency (in Hz)?
16. A signal $x[n]$ has a period of 100 milliseconds. What is its frequency in Hz?
17. What is the distance in Hz between the frequency bins of the $N = 50$ point Discrete Fourier Transform of a signal with a sample rate of 10,000 Hz?

18. An STFT $X[m, k]$ is computed from an audio signal with a sample rate of 5000 Hz, using a DFT of size 500 samples and a hop size of 250 samples. The STFT bin $X[2000, 100]$ corresponds to what point in time (in seconds)?

19. An audio signal has a sample rate of 48 kHz. An onset novelty function is computed with a hop size of 2000 samples. What is the sample rate of the novelty function?

20. What is the nyquist frequency of $fs = 22050Hz$?

21. What is the zero crossing rate for a signal that crosses zero 200 times in a frame length of 2000 samples?

22. What is the resulting shape of a feature set (n_samples, n_features) that has a feature vector size of 10 extracted from 30 seconds of audio with a sample rate of 10,000, using a hop size of 500?

Implementation

Using the files provided in **midterm.zip**, answer *one* of the two questions in this section. Your answer should include both descriptions and code (python). Any libraries and references you use should be clearly noted. You will be assessed on the thoughtfulness and completeness of your response. You will not be graded solely on the correctness of the output, though it will be considered. (8 points)

23. Provide a detailed description of note onset detection with spectral flux. Using the file **onset.wav**, describe the process step-by-step, from raw input to extracted onset times.

    Your response should include the following: **(a)** Briefly describe the goal of note onset detection, **(b)** Plot a spectrogram of the audio, correctly labeled, and describe what parameters you used and why you chose those parameters. **(c)** Describe the spectral flux calculation without using the equation. **(d)** Describe the process of peak-picking as it relates to the novelty function and note onsets. Describe the parameters you chose and your reasoning. **(e)** Plot the output novelty function, correctly labeled, with the selected peaks marked.

24. Implement a 2-class k-Nearest Neighbor instrument classification experiment. The file **knn_train.wav** contains exactly 40 seconds of trombone followed by exactly 40 seconds of trumpet. The file **knn_test.wav** contains exactly 15 seconds of trombone followed by exactly 15 seconds of trumpet. Extract MFCCs from both files and label the first half as trombone and the second half as trumpet. Then run k-NN classification.

    Your response should include the following: **(a)** Plot a mel-frequency spectrogram of the training set, correctly labeled. **(b)** Explain what an MFCC is, what information it captures, and how it is calculated (equations and specific steps are not required). Discuss the parameters you chose for MFCC extraction. **(c)** Create two datasets and labels, one for training and one for testing, as indicated in the file name. Describe the purpose of a training and test set. **(d)** Run the experiment on at least 2 values for k, and choose the best one. Explain why you made that choice. **(e)** Print the confusion matrix, which does not need to be normalized. Describe the results.