

## Problem 1

### 1. Speech Emotion Recognition

The basic goal of speech emotion recognition is to understand and retrieve desired emotions from speech audios, which is to recognize the underlying emotional state of a speaker from her voice. Compared to other types of biological signals used in emotion recognition, such as electrocardiogram, speech signals usually can be acquired more readily and economically. There are many applications of detecting the emotion of the persons like in the interface with robots, audio surveillance, clinical studies, etc.

### 2. Source Separation

Source separation is to separate a set of source signals from mixed signals without or with little aid of extra information. The most well-known source separation situation is the cocktail party problem, where there is a group of people talking at the same time and you need to isolate the speech of a single person.

### 3. Automatic Music Transcription

Automatic music transcription is the design of computational algorithms to convert acoustic music signals into some form of music notation. This is a challenging task because it comprises several subtasks, including multi-pitch estimation, onset and offset detection, instrument recognition, beat and rhythm tracks, interpretation of expressive timing and dynamics, and score typesetting.

### 4. Genre Classification

Genre Classification is to classify music into different genres. Automatic music genre classification is a very interesting problem in the context of MIR because it enables systems to perform content-based music recommendation, organizing musical databases and discovering media collections.

### 5. Music Detection

Music detection refers to the task of finding music segments in an audio file. The two main applications of music detection algorithms are (1) the automatic indexing and retrieving of auditory information based on its audio content, and (2) the monitoring of music for copyright management. Additionally, the detection of music can be applied as an intermediate step to improve the performance of algorithms designed for other purposes.

### 6. Audio Classification

Audio signal classification system analyzes the input audio signal and creates a label that describes the signal at the output. These are used to characterize both music and speech signals. The categorization can be done based on pitch, music content, music tempo, and

rhythm. Audio classification can be used for audio scene understanding which in turn is important so that an artificial agent can understand and better interact with its environment.

## 7. Composer Identification

Automatic composer-identification is a complex task that remains a challenge in the field of music information retrieval. There are two problems here, first, that of constructing an automatic system that can distinguish between music written by different composers; and, second, that of identifying the musical properties that are important for this task. The latter can offer interesting insights for music theorists.

## 8. Automatic Lyrics-to-Audio Alignment

The task of automatic lyrics-to-audio alignment has as an end goal the synchronization between an audio recording of singing and its corresponding written lyrics. The beginning timestamps of lyrics units can be estimated on different granularity: phonemes, words, lyrics lines, phrases. For this task word-level alignment is required.

## 9. Audio Beat Tracking

The automatic beat tracking task aims to track each beat locations in a collection of sound files. The algorithms will be evaluated in terms of their accuracy in predicting beat locations annotated by a group of listeners.

## 10. Acoustic Monitoring

Many animals produce sound for communication and navigation. So ecologists and conservation researchers use sound to study wildlife, from identifying birdsong in the field, to surveying bats with handheld detectors, to using state-of-the-art remote sound recorders to monitor the sounds of animals and ecosystems.

**Problem 2**

1. **Codes:** problem2.m, 2-a

**File:** myvoice.wav.

2. **Codes:** problem2.m, 2-b **File:** myvoice\_noisy.wav

File	RMS Value	Code
myvoice.wav	0.029910321812699	problem2.m, 2-b, Line 10
niose.wav	0.052470234458801	problem2.m, 2-b, Line 15
myvoice_noisy.wav	0.093306744410803	problem2.m, 2-b, Line 20

3. **Codes:** problem2.m, 2-c

**File:** magnitude spectrum: 2-c.jpg

**Questions:**

**How long is this window in milliseconds?**

Window Length/fs = 512/16k = 32ms

**What is the (real) frequency resolution?**

Calculate: 16k/512 = 31.25

Codes located at problem2.m, 2-c, Line 35-36

4. **Codes:** problem2.m, 2-d

**File:** spectrogram: 2-d.jpg

5. **Codes:** problem2.m, 2-e

**File:** myvoice\_telephone.wav; myvoice\_telephone\_zerophase.wav

### Problem 3

1. **Codes:** problem3.m, 3-a

**File:** complextone.wav

**MIDI number:** [45 57 69 81 105], codes located at line 18, 21

2. **Codes:** problem3.m, 3-b

**File:** complextone\_Gaussian.wav

3. **Codes:** problem3.m, 3-c

**File:** Shepardtone.wav