

# Part1

The following is the scenario of each theory question type appearing in the test in 2023 and 2024.

20230623

Q2 time-scale ✓

Q3 a delay line ✓

20230112

Q2 RIR ✓

Q3 time-pitch-scale ✓

20230911

Q2 time-pitch-scale ✓

Q3 RIR ✓

20230711

Q2 a delay line ✓

Q3 synthesizing a reverberated audio signal – RIR ✓

20230206

Q2 synthesizing audio signals – Wavetable synthesis method ✓

Q3 classifying a set of sounds into two different classes - a supervised machine learning approach

20231104

Q2 window function, frequency resolution and accuracy. ✓

Q3 delay line ✓

20240122

Q2 synthesizing a reverberated audio signal – RIR ✓

Q3 synthesizing audio signals – wavetable, granular, additive ✓

20240617

Q2 window function + frequency resolution and accuracy ✓

Q3 audio reverberation. – RIR ✓

20240212

Q2 delay lines ✓

Q3 time and pitch scaling ✓

20240708

Q2 a delay line ✓

Q3 synthesizing a reverberated audio signal – RIP ✓

## Part2

Here is a summary of the frequency of each theoretical topic:

### 1. Delay Line

- Frequency: 6 times
- Dates: 20230623 (Q3), 20230711 (Q2), 20231104 (Q3), 20240212 (Q2), 20240708 (Q2)

### 2. Synthesizing a Reverberated Audio Signal – RIP(Room Impulse Response)

- Frequency: 6 times
- Dates: 20230711 (Q3), 20240122 (Q2), 20240617 (Q3), 20240708 (Q3), 20230112 (Q2), 20230911 (Q3)

### 3. Time-Pitch-Scale

- Frequency: 4 times
- Dates: 20230112 (Q3), 20230911 (Q2), 20230623 (Q2), 20240212 (Q3),

### 4. Window Function, Frequency Resolution and Accuracy

- Frequency: 2 times
- Dates: 20231104 (Q2), 20240617 (Q2)

### 5. Synthesizing Audio Signals – Wavetable, Granular, Additive, Wavetable Synthesis Method

- Frequency: 2 time
- Date: 20240122 (Q3), 20230206 (Q2)

### 6. Classifying a Set of Sounds into Two Different Classes - A Supervised Machine Learning Approach

- Frequency: 1 time
- Date: 20230206 (Q3)

Summary:

- **Delay Line** is the most frequently appearing topic, occurring 6 times.
- **Synthesizing a Reverberated Audio Signal – RIP** is the same, appearing 6 times.
- **Time-Pitch-Scale** appears 4 times, making it another common topic.
- Other topics appear less frequently, mostly 1-2 times.

## Part 3

**Answer the questions in the corresponding paper. You can check it when you do the topic on the theory part.**

### 1. Delay line

Q1. A delay line shifts an audio signal in time by storing samples and releasing them later. It stores the input signal in a buffer and outputs it after a specified amount of time has passed. Formally, if  $x(n)$  is the input, then the output is  $y(n)=x(n-D)$ , where  $D$  is the delay in samples. It creates echoes or timing effects commonly used in audio processing.

Q2. An integer delay line shifts the signal by a whole number of samples, offering simple, discrete timing. A fractional delay line enables sub-sample (non-integer) delays, requiring interpolation to estimate values between samples. A time-variant fractional delay line goes further by allowing the fractional delay amount to change over time, enabling dynamic effects like vibrato or pitch modulation.

Q3. Fractional and time-variant fractional delay lines need an interpolator because the required delay often falls between sample points, they require accessing sample values at non-integer positions. Without interpolation, you can only shift by whole samples. Interpolation fills in these fractional points smoothly, ensuring smooth, accurate fractional delays. Integer delay lines do not need it, as they only shift by whole samples and directly read from the stored buffer without fractional offsets.

Q? A time-variant fractional delay line provides a delay that changes over time and is not restricted to integer sample steps. By allowing fractional sample delays, it can smoothly modulate the timing of a signal, producing effects like vibrato, chorus, or pitch variations without abrupt temporal jumps.

### 2. Time scaling

Q1. Time scaling changes how long an audio signal plays without altering its original pitch. If you make it slower, the signal's duration increases, and if you speed it up, the duration shortens. It's like stretching or compressing the timeline while keeping the same harmonic relationships among frequencies. original frequencies remain unchanged

From PPT:

Time Scaling Controlling and modifying the time-evolution of a signal not altering spectral components. To slow down or speed up a given signal, possibly in a time-varying manner, without altering the signal's spectral content

Q2. Pitch scaling alters the frequencies in an audio signal without changing its overall duration. Raising the pitch moves every frequency upward, producing a higher-sounding result, while lowering it shifts frequencies downward. Unlike time scaling, the tempo remains the same, but the musical key or tonal range changes. the time length remains the same.

From PPT:

Pitch Scaling can Control and modify the pitch of a signal and not alter the temporal evolution, To modify the pitch of the signal, possibly in a time-varying manner, without altering the signal's time-evolution.

Q3. The duality principle suggests that time and pitch scaling are inversely related. Stretching the signal in time corresponds to lowering the perceived pitch if done by basic resampling, and compressing it raises the pitch. Advanced algorithms can independently control these two aspects, but they remain intertwined at a fundamental level.

From PPT:

Duality principle Definition: Time scaling can be obtained combining pitch scaling and time warping. Pitch scaling can be obtained combining time scaling and time warping

Q4. PSOLA is a technique used on speech signals to perform time and pitch scaling. It segments the signal at pitch periods, then overlaps and adds these segments to modify pitch or duration. By changing the overlap rate, it can scale time; by shifting segments in time, it can scale pitch. Thus, PSOLA applies both time and pitch scaling.

From PPT:

PSOLA Principle: The length of the repeated/discarded segments is adjusted according to the local value of the pitch given by a preliminary pitch estimation

Q5 A two-second sinusoid at 440 Hz contains 880 cycles over its duration. Applying time-scaling with factor 2 doubles the length to four seconds, effectively halving its frequency to 220 Hz. The result is a slower, lower-pitched tone, as the sinusoid now takes twice as long to complete each cycle.

Q6. Pure time-stretching or resampling also changes pitch because it literally plays the samples faster or slower. In contrast, time-scaling techniques like phase vocoding or PSOLA modify the waveform structure to keep pitch constant while adjusting duration. Hence, resampling affects both speed and pitch, whereas time-scaling affects duration only.

**3. RIP:**

Q1. A Room Impulse Response (RIR) captures how an impulsive sound from a source travels to a receiver in a room. It encompasses the direct sound, early reflections, and the reverberant tail caused by multiple higher-order reflections. Essentially, it characterizes the room's acoustic fingerprint for that specific source–receiver setup.

Definition: Transfer function between a sound source and a receiver within an environment. It captures information about all audio reflections in the environment. We assume that audio propagation can be modeled as a linear filtering operation. The RIR is a unique acoustic fingerprint, helping recreate the room's sonic character when applied to dry signals

Q2. A typical RIR can be split into three main parts: Direct Sound: A strong initial peak. Early Reflections: A few discrete echoes arriving soon afterward. Reverberant Tail: A longer, dense decay formed by overlapping reflections that gradually diminish in amplitude.

Q3. Definition: Time needed by the EDC to decrease by 60dB  
Reverberation time  $T_{60}$  is formally defined as the time required for the sound energy in a room to drop by 60 dB after the source stops. A longer  $T_{60}$  means a more “echoey” space, while a shorter  $T_{60}$  indicates a more controlled environment.

Q4. To apply reverberation, convolve the dry signal with the RIR. Convolution blends the original sound with the room's impulse response, adding realistic reflections and decay.  $y(n)=x(n)*h(n)$ . where  $x(n)$  is the dry audio and  $h(n)$  is the RIR. This convolution simulates the room's acoustic response, adding realistic reflections and decay to the original sound.

#### 4. Wavetable synthesis:

Q1. Wavetable synthesis stores one period of a waveform in a table. By reading through it at varying speeds, you produce different pitches without recreating the wave each time. This technique saves computation and provides consistent timbre, making it efficient and popular in many synthesizers and digital instruments.

Q2. Pros include efficient CPU usage, stable tuning, and a wide variety of sounds from a single table.

Cons include limited timbral flexibility if the stored waveform is static, and potential aliasing or unnatural sounds if the playback rate differs greatly from the original recording's fundamental frequency.

Q3. Granular synthesis splits audio into tiny “grains” of milliseconds. By rearranging, overlapping, and modulating these grains, it creates evolving textures, time-stretching,

and pitch-shifting effects. This method offers great flexibility and can produce complex, organic sounds that traditional synthesis methods find hard to achieve.

Q4. Additive synthesis builds complex sounds by summing multiple sine waves at different frequencies, amplitudes, and phases. Each sine represents a partial or harmonic of the final tone. By adjusting these parameters over time, you can sculpt evolving timbres and replicate many naturally occurring sonic phenomena in a mathematically straightforward way.

## 5. Window functions:

Q1. Window functions limit the time-domain signal and affect the frequency-domain representation, the window function causes spectral leakage and broadening. To resolve closely spaced sinusoids, you need a sufficient frequency separation (at least one main lobe width apart). A longer window improves frequency resolution but reduces time resolution, and some leakage or amplitude bias can remain.

Q2. Frequency resolution concerns the smallest frequency difference you can separate in the spectrum, often tied to window (frame) length. Accuracy is about how close the estimated frequency is to its true value. High resolution does not guarantee high accuracy; you may pinpoint two peaks but still estimate their exact frequencies imperfectly.